(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number
WO 01/27158 A2

(51) International Patent Classification⁷: C07K 14/705

(21) International Application Number: PCT/US00/27582

(22) International Filing Date: 6 October 2000 (06.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/158,615 8 October 1999 (08.10.1999) US
60/184,809 24 February 2000 (24.02.2000) US

(71) Applicants *(for all designated States except US)*: **DIGISCENTS** [US/US]; Suite 720, 1814 Franklin Street, Oakland, CA 94612 (US). **YEDA RESEARCH AND DEVELOPMENT CO., LTD.** [IL/IL]; Weizmann Institute of Science, P.O. Box 95, 76100 Rehovot (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BELLENSON, Joel** [US/US]; 244 Lakeside Drive, Apartment 15, Oakland, CA 94612 (US). **SMITH, Dexster** [US/US]; 868 Trestle Glen Road, Oakland, CA 94610 (US). **LANCET, Doron** [IL/IL]; 15 Weizmann Street, 76280 Rehovot (IL). **GLUSMAN, Gustavo** [IL/IL]; 33/37 Ha'Alon Street, 79845 Bnei Ayish (IL). **FUCHS, Tania** [IL/IL]; 12 Harav neria Street, 75751 Rishon LeZion (IL). **YANAI, Itai** [US/US]; 55 Leicester Street, Brookline, MA 02146 (US).

(74) Agents: **CERPA, Robert, K.** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/27158 A2

(54) Title: OLFACTORY RECEPTOR SEQUENCES

(57) **Abstract:** The present invention provides polynucleotide sequences which encode polypeptides involved in olfactory sensation. The present invention also provides the polypeptides encoded by these polynucleotide sequences, vectors comprising these polynucleotide sequences and host cells transfected with these polynucleotide sequences. The present invention further provides for functional variants and homologues of these polynucleotide sequences and the polypeptides encoded by these polynucleotides. Libraries of polypeptides are also provided. Also included in the present invention is the use of these polypeptides and libraries of polypeptides in screening odorant molecules to determine the correspondence (scent representation, scent fingerprint or scent profile) between individual odorant receptors (the polypeptides) and particular odorant molecules. Also encompassed by the present invention is the use of the scent representation, scent fingerprint or scent profile to re-create and edit scents.

OLFACTORY RECEPTOR SEQUENCES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority benefit of United States Provisional Patent Application Serial No. 60/158,615, filed on October 8, 1999, and United States Provisional Patent Application Serial No. 60/184,809, filed on February 24, 2000. The contents of those applications are hereby incorporated by reference herein in their entirety.

STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH

Not applicable.

TECHNICAL FIELD

The present invention is in the field of human olfactory receptors and their use in screening for olfactory agonists and antagonists. The present invention pertains to isolated nucleotide sequences which encode human olfactory receptors and also to the proteins encoded by said nucleotide sequences. The present invention also encompasses vectors comprising the nucleotide sequences of the invention and further, host cells transfected with said vectors. The present invention also allows for the determination of primary scents and the identification of the odor receptors which are encoded to detect these primary scents as well as the determination of secondary scents and the identification of combinations of odor receptors which are encoded to detect such secondary scents.

BACKGROUND ART

Our sense of smell plays an important role not only in our appreciation of our surroundings such as the smell of flowers or new mown grass, but also evolved as a survival skill. Numerous odorant molecules can be detected at extremely low concentrations, providing early warning of danger, such as the smell of smoke or contaminated food. Indeed, a potent example of this is that most pregnant women experience a heightened sense of smell, presumably to protect the fetus from the deleterious effects of food poisoning.

It is estimated that humans can detect millions of different molecular species; however, our nose can discriminate only a fraction of these different chemicals (Mombaerts *Curr. Opin. Genet. Dev.* 1999 **9**, 315-320), usually estimated at about 10,000 odorants (Axel, *Scientific American* 1995, **October**, 154-159). Odorants for terrestrial species such as humans, are volatile (air born) ligands which are detected by the olfactory system. Odorants have vastly different chemical structures and subtle differences can lead to pronounced changes in the perceived odor (Mombaerts, *supra*). For instance, when the hydroxyl group of octanol is replaced by a carboxyl group to give octanoic acid, its perceived odor changes from orange and rose-like to rancid and sweaty (Malnic *et al.*, *Cell* 1999 **96**, 713-723). The basis for these feats of sensory perception are just beginning to be understood at a cellular and molecular level.

The olfactory system contains millions of olfactory sensory neurons (OSNs) located in the olfactory epithelium of the nasal cavity. In humans, the olfactory epithelium occupies an area of approximately 5 cm^2. The OSNs are bipolar with one end extending through the supporting cell into the mucosal layer, terminating in hairlike cilia. These cilia are the site of the olfactory receptors (OR) where the odorant ligands are thought to bind (Mombaerts *Curr. Opin. Genet. Dev.* 1999 **9**, 315-320, Hildebrand *et al., Annu. Rev. Neurosci.,* 1997, **20**, 595-631). The OSNs also have a single unbranched axon which leads to the olfactory bulb, a part of the brain containing approximately 2000 glomeruli where the axons terminate and initial processing of the sensory code takes place. OSNs expressing the same OR are randomly interspersed throughout the olfactory epithelium, but in both the nose and the bulb, information derived from different ORs is strictly segregated; each OSN in the nose and each glomerulus in the olfactory bulb appear to be dedicated to input from one or few OR type(s) (Malnic *et al.*, *Cell* 1999 **96**, 713-723). It also appears that the location of the glomeruli are conserved across individuals of a species, providing the first spatial processing of particular odorant patterns (Mombaerts *Curr. Opin. Genet. Dev.* 1999 9, 315-320). The domains in the olfactory bulb for

different odors may overlap, but the overall patterns are distinct (Hildebrand *et al.*, *supra*), therefore, it should be possible to identify and reproduce the characteristic pattern of a given odorant. Output neurons project from the olfactory bulb to the primary olfactory cortex and from there to the higher cortical areas of the brain and to the limbic system (Malnic *et al.*, *supra*; Hildebrand *et al.*, *supra*, **20**, 595-631).

Until the identification of a large family of genes encoding putative odorant receptors (Buck & Axel *Cell* 1991 **65**, 175-187), progress towards understanding the process of odor recognition was negligible. In recent years there has been an explosion in this field as more and more putative odor receptors are isolated and cloned. The odorant receptor gene products have thus far been characterized through homology as seven transmembrane domain G protein-coupled receptors (GPCR). It is estimated that there are probably 500-750 OR-like sequences in humans, while there are 500-1000 OR genes in rat and mouse (Mombaerts *Curr. Opin. Genet. Dev.* 1999 **9**, 315-320). In mice, OR-like sequences make up approximately 1% of their genome, the largest known family in the mammalian genome, surpassing the complexity of even the immunoglobin and T-cell antigen receptor gene families (Mombaerts, *supra*). The OR are concentrated on the surface of the OSN's mucus coated cilia and it is thought that odorant molecules bind to the OR in the olfactory epithelium and thereby initiate signal transduction. Current interpretation of recent experimental evidence favors the idea that each neuron expresses only one, or very few, ORs. Since mammals can detect at least 10,000 odors and there are approximately 1,000 or fewer ORs, each of the ORs must respond to several odorant molecules, and each odorant molecule must bind to several receptors. It is believed that various receptors respond to discrete parts of an odorant molecule's structure and that an odorant consists of several chemical groups each of which bind a characteristic receptor (Axel *Scientific American* 1995, **October**, 154-159; Malnic *et al.*, *Cell* 1999 **96**, 713-723).

The main signal transduction pathway mediated by OR homologues in vertebrate species involves G protein-mediated stimulation of adenylyl cyclase activity, resulting in cAMP elevation that opens cyclic-nucleotide gated channels with a non-specific cation selectivity (Mombaerts *Curr. Opin. Genet. Dev.* 1999 **9**, 315-320). However, there are still numerous unanswered questions and recently it has come to light that 38-76% of the human gene OR sequences that are being reported may be pseudogenes and therefore incapable of expressing the proteins that encode the olfactory receptors. Some of the incidences may be due to the method of extracting the genomic DNA libraries (Mombaerts, *supra*). Few pseudogenes have been found in other vertebrates and their incidence in libraries from testicular DNA is also

rare (Hildebrand *et al.*, *Annu. Rev. Neurosci.*, 1997, **20**, 595-631). cDNA should not contain pseudogenes. There are a number of examples of ORs which have been successfully expressed and reactions to certain odorant ligands have been determined (Malnic *et al.*, *Cell* 1999 **96**, 713-723; Mombaerts, *supra*; Zhao *et al.*, *Science* 1998 **279**, 237-242).

Some attempts to express the ORs in heterologous cell lines resulted in the formation of inclusion bodies rather than the insertion of the proteins into the membrane (Kiefer *et al.*, *infra*). However, purification of the receptors after expression in *E. coli* and their insertion into lipid vesicles facilitates the use of these receptors in odorant ligand screening using a combination of photoaffinity labeling and Trp fluorescence (Kiefer *et al.*, *Biochemistry* 1996 **35**, 16077-16084). In addition, a functional human OR receptor protein has been expressed in HEK-293 cells and oocytes and found to interact with odorant ligands (Wetzel et al., *J. Neurosci.* 1999 **19**, 7426-7433). There have also been, a number of successful efforts of expressing cDNA in insect Sf9 cells using *baculovirus* vectors (Mombaerts *Annu. Rev. Neuorsci.* 1999) as well as assays with neuronal tissue (Malnic *et al.*, *Cell* 1999 **96**, 713-723; Zhao et al., 1998; Firestein *et al.*, WO 98/50081). In addition, recent work accomplished the expression of chimeric mouse olfactory receptor sequences in HEK-293 cells and showed their reactivity towards a panel of odorant ligands, some at micromolar concentrations (Krautwurst *et al.*, *Cell* 1998 **95** 917-926). The drawback to expression in heterologous cell systems is the lack of working signal transduction pathways which can be used to detect responses to odorant ligands; these drawbacks can be overcome with methods known in the art (e. g. U.S. Pat. No. 5,798, 275). There are also methods of expressing and assaying functional neuronal receptors in neuronal cells, including methods for detecting particular odorant ligand specificity (Malnic *et al.*, *supra*; Zhao, *supra*; Firestein *et al, supra*).

Other publications of interest are: Chemical Senses 6: 343-349 (1981); Proc. Natl. Acad. Sci. USA 79: 670-674 (1982); Proc. Natl. Acad. Sci. USA 81(6): 1859-1863 (1984); Nature 316: 255-258 (1985); Brain Research 368: 329-338 (1986); J. Biol. Chem. 261: 1299-1305 (1986); Proc. Natl. Acad. Sci. USA 83(13): 4947-4951 (1986); J. Neurosci. 6: 2146-2154 (1986); J. Neurochem. 47: 1527-1533 (1986); Chemical Senses 13: 191-204 (1988); Biochem. J. 260:121-126 (1989); J. Biol Chem. 264: 6780-6785 (1989); Biochim. Biophys. Acta 1013: 68-72 (1989); J. Biol. Chem. 264: 18803-18807 (1989); Biochemistry 29: 7433-7440 (1990); FEBS lett. 270: 24-29 (1990); Chemical Senses 15: 529-536 (1990); Eur. J. Biochem. 196: 51-58 (1991); Nature 349: 790-793 (1991); Neurosci. Lett. 141: 115-

118 (1992); Developmental Brain Res. 73: 7-16 (1993); Proc. Natl. Acad. Sci., USA 90: 3715-3719 (1993); Human Mol. Genetics 3: 229-235 (1994); Eur. J. Biochem. 225: 1157-1168 (1994); European Journal of Biochemistry 238: 28-37 (1996); Receptors and Channels 4: 141-147 (1996); Genomics 37(2): 147-160 (1996); Protein Science 8: 969-977 (1999); Genomics 53: 56-68 (1998); Genomics 61:24-36 (1999); Genomics 63: 227-245 (2000); Trends in Neurosci. 7:35-36 (1984); Ann. Rev. Neurosci. 9:329-355 (1986); Trends Biochem. Sci. 12:63-66 (1987); Nature 351: 275-276 (1991); Nature 353: 799-800 (1991); Current Biol. 3(10): 668-674 (1993); Nature 372:321-322 (1994); Essays in Biochemistry. 33: 93-104 (1998); and Nature, 398 (6725): 285-287 (1999).

However, despite the forgoing, there has been relatively little work with human olfactory receptors, in particular in determining the sequences of large numbers of receptors, and less progress in determining the correspondence between particular human olfactory receptors and the scent(s) to which they respond.

All publications cited herein are hereby incorporated by reference in their entirety.

DISCLOSURE OF THE INVENTION

An object of the invention is to determine the correspondence between ORs and the scent(s) to which they respond. Once this is accomplished, scents can be both analyzed and re-created for enhancing human experiences or eliciting particular responses. The present invention pertains to isolated polynucleotide sequences encoding polypeptides involved in olfactory sensation. The present invention also pertains to the proteins encoded by said nucleotide sequences. The present invention also encompasses vectors comprising the nucleotide sequences of the invention and further, host cells transfected with said vectors. The present invention also allows for the determination of primary scents and the identification of the odor receptors which are encoded to detect these primary scents as well as the determination of receptor complex scent components and the identification of combinations of odor receptors which are encoded to detect such receptor complex scent components scents.

The invention provides isolated polynucleotide sequences encoding polypeptides involved in olfactory sensation that are isolated from human olfactory epithelial tissue. The invention further provides expression vectors containing such nucleotide sequences. Also provided by the invention are purified polypeptides encoded by the nucleotide sequences. The invention further provides transformed cells which comprise a suitable host cell transfected with a suitable expression vector containing the nucleotide sequence encoding the receptor. The present invention also encompasses nucleotide sequences isolated from human olfactory epithelial tissue which encode receptors capable of binding odorant molecules. The invention further provides expression vectors containing such nucleotide sequences and homologues of both the polynucleotides and polypeptides. Further, the invention provides a means of using the nucleotide sequences of the invention in a method of screening odorant ligands to determine the specific binding of odorant molecules to a particular receptors, and further, determining the component odorant molecules of subjectively experienced smells, determining the combination odorant molecules and receptor stimulation or inhibition to re-create a particular scent. The binding of odorant molecules by the receptors encompassed in the present invention includes binding resulting in both the agonism (excitation/activation) and antagonism (inhibition/blocking) of receptor function(s) upon binding of the molecule.

Accordingly, the invention includes an isolated polynucleotide comprising a sequence encoding a polypeptide which is involved in olfactory sensation. The OR polypeptides encoded are found within the sequences depicted in polynucleotide sequences SEQ ID NO:1 through SEQ ID NO: 73 and SEQ ID NO:111 through SEQ ID NO:152, or a nucleotide sequence at least 95% homologous to said sequences. The invention also encompasses the translation products of those sequences. The invention further comprises expression vectors comprising said sequences, host cells containing such expression vectors and/or expressing the polypeptide encoded therein, or phage displaying the polypeptide encoded by the sequences. The use of functional fragments of receptors is also encompassed by the invention. Preparations of receptors, further including biological or synthetic molecules which maintain the stability and functional structure of the receptors, are also included in the invention. The invention further encompasses fragments of said polynucleotides which can be used as probes or primers to identify additional polynucleotide sequences through techniques known in the art, including those fragments depicted in SEQ ID NOs: 74-105.

The invention also includes additional isolated polynucleotide comprising a sequence encoding a polypeptide which is involved in olfactory sensation. The OR polypeptides

encoded are found within the sequences depicted in polynucleotide sequences SEQ ID NO:153 through SEQ ID NO: 1084, or a nucleotide sequence at least 95% homologous to said sequences. The invention also comprises the translation products of those sequences. The invention further comprises expression vectors comprising said sequences, host cells containing such expression vectors and/or expressing the polypeptide encoded therein, or phage displaying the polypeptide encoded by the sequences. The use of functional fragments of receptors is also encompassed by the invention. Preparations of receptors, further including biological or synthetic molecules which maintain the stability and functional structure of the receptors, are also included in the invention.

The invention also encompasses an isolated and purified olfactory receptor polypeptide scomprising the sequence of SEQ ID NO: 1085 through SEQ ID NO: 2008, or a polypeptide sequence that is at least about 95% homologous to a polypeptide sequence of the group consisting of SEQ ID NO: 1085 through SEQ ID NO: 2008 and having olfactory receptor function. Host cells expressing such polypeptides and phages displaying such polypeptides are also encompassed by the invention. The use of functional fragments of receptors is also encompassed by the invention. Preparations of receptors, further including biological or synthetic molecules which maintain the stability and functional structure of the receptors, are also included in the invention.

Scents can be captured, analyzed and recorded by a sensory device using various methods. Scent capture can be initiated by the user or by an automatic sensing system. A scent can be analyzed in terms of its interaction with olfactory neurons of a mammalian, preferably human, olfactory system, or by the expression of individual receptors under appropriate conditions and appropriate assay conditions in multiwell plates or in terms of its perception by a panel of mammalian, preferably human, subjects. The interaction with olfactory neurons can be determined experimentally, in vitro, by determining the interaction of an odorant with olfactory receptors of a given type. Alternatively, the interaction with olfactory receptor can be determined using a computer simulation which provides information regarding the interaction of an odorant with the olfactory receptors. A panel of subjects can be used to represent odors in terms of their perception. The data so generated can be used to represent a scent in a manner which can be recorded in digital or other format, stored in media such as computer memory, disks, or printed format, and transmitted over a data network. The representation of the scent can be used to re-create the scent at a local or remote site using an emitter module. The

representation of the scent allows for scent editing, where desirable aspects of an odor are enhanced or added and undesirable aspects are attenuated or eliminated.

Accordingly, the invention also embraces libraries of olfactory receptors suitable for determining the interaction pattern of a composition with the receptors, comprising the expression products of at least two polynucleotides of SEQ ID NO:1 through SEQ ID NO: 73, SEQ ID NO:111 through SEQ ID NO:152, and SEQ ID NO: 153 through SEQ ID NO: 1084, where the polynucleotides encode functional olfactory receptors; or functional fragments of the expression products. Libraries of at least 50, 100, 200, or 500 receptors are also encompassed by the invention.

Also encompassed by the invention are libraries of olfactory receptors suitable for determining the interaction pattern of a composition with the receptors, comprising at least two polypeptides of SEQ ID NO: 1085 through SEQ ID NO: 2008, where the polypeptides are functional olfactory receptors; or functional fragments of the polypeptides. Libraries of at least 50, 100, 200, or 500 receptors are also encompassed by the invention.

The invention also embraces methods for determining the binding pattern of a composition with olfactory receptors, involving exposing the composition to an olfactory receptor library, and determining whether the composition binds to each olfactory receptor, thereby determining the overall binding patter of the composition. In additional embodiments, the method also involves determining the approximate binding constant with which the composition, or the various chemicals within the composition, bind to the receptors; determining whether a receptor or functional fragment thereof is activated; and determining the absolute amount of activation, or amount of activation relative to another receptor or a control substance. The composition can consist essentially of one compound or chemical, or can comprise at least two compounds or chemicals.

The invention also embraces DNA arrays or DNA chips comprising the DNA segments derived from any combination of, or each of, SEQ ID NO: 153 through SEQ ID NO: 1084. The invention also embraces a method of determining differences among one or more individuals with respect to their olfactory faculties, comprising the steps of comparing the olfactory DNA of each individual against the array or chip.

The invention also embraces a method to determine single nucleotide polymorphisms in olfactory receptors, comprising the steps of uniquely amplifying olfactory receptor sequences from DNA obtained from one or more individuals, based on

primers designed according to the first 25 bases and the last 25 bases of any combination of, or each of, SEQ ID NO: 153 through SEQ ID NO: 1084, and determining the similarities and differences between said amplified DNA and the corresponding receptor from SEQ ID NO: 153 through SEQ ID NO: 1084.

Brief Description of the Drawings

Figure 1 depicts the isolated polynucleotide sequences, which encode polypeptides involved in olfactory sensation, corresponding to SEQ ID NOs: 1 - 73.
Figure 2 depicts the isolated polynucleotide sequences, which encode polypeptides involved in olfactory sensation, corresponding to SEQ ID NOs: 111 - 152.

Detailed Description of the Invention

The present invention provides isolated polynucleotides comprising sequences that encode polypeptides which are involved in olfactory sensation and which can be used to screen odorant ligands, *e.g.*, odorant receptor agonists and antagonists.

Definitions

The term "olfactory receptor" (OR) refers to a polypeptide involved in olfactory sensation. An "olfactory receptor polynucleotide" or "OR polynucleotide" is a polynucleotide encoding a polypeptide involved in olfactory sensation.

The term "odorant ligand" as employed herein refers to a molecule that has the potential to bind to an olfactory receptor. Equivalent terms employed herein include "odorant", "odorant molecule" and "odorant compound". The term "binding" or "interaction" as used herein with respect to odorant ligands refers to the interaction of ligands with the receptor polypeptide where the ligands may serve as either agonists and/or antagonists of a given receptor or receptor function. An odorant ligand may thus directly cause a perception of odor (an agonist), or may block the perception of odor (an antagonist). An odorant ligand may include, but is not limited to, molecules which interact with polypeptides involved in olfactory

sensation. Odorant ligands and molecules which interact with olfactory receptors are generally small, approximately 1000 Daltons, more preferably approximately 750 Daltons, more preferably approximately 500 Daltons, or even more preferably approximately 300 Daltons, hydrophobic molecules with a variety of functional groups. Small changes in structure can induce profound changes in odorant ligand binding and hence in the odor perceived by an individual.

A more detailed description of these sequences, as well as how these sequences were obtained, is provided below.

As used herein, a "polynucleotide" is a polymeric form of nucleotides of any length, which contain deoxyribonucleotides, ribonucleotides, and/or their analogs. The terms "polynucleotide", "nucleotide" and "nucleic acid" as used herein are used interchangeably. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The term "polynucleotide" includes double- , single-stranded, and triple-helical molecules. Unless otherwise specified or required, any embodiment of the invention described herein that is a polynucleotide encompasses both the double-stranded form and each of two complementary single-stranded forms known or predicted to make up the double stranded form. Not all linkages in a polynucleotide need be identical.

The following are non-limiting examples of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, primers, and adaptors. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. The use of uracil as a substitute for thymine in a deoxyribonucleic acid is also considered an analogous form of pyrimidine.

In the context of polynucleotides, a "linear sequence" or a "sequence" is an order of nucleotides in a polynucleotide in a 5' to 3' direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polynucleotide. A "partial sequence" is a linear sequence of part of a polynucleotide which is known to comprise additional residues in one or both directions.

If present, modification to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by

conjugation with a labeling component. Other types of modifications included in this definition are, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, cabamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., α-anomeric nucleic acids, peptide nucleic acids, etc.), as well as unmodified forms of the polynucleotide(s).

Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid supports. The 5' and 3' terminal OH groups can be phosphorylated or substituted with amines or organic capping group moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups.

Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, but not limited to, 2'-O-methyl-, 2'-O-allyl, 2'-fluoro- or 2'-azido-ribose, carboxcyclic sugar analogs, α-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs such as methyl riboside.

Although conventional sugars and bases will be used in applying the method of the invention, substitution of analogous forms of sugars, purines and pyrimidines can be advantageous in designing a final product, as can alternative backbone structures like a polyamide backbone such as those used in peptide nucleic acids (PNAs).

A polynucleotide or polynucleotide region has a certain percentage (for example, 75%, 80%, 85%, 90%, 95% or 99%) of "sequence identity" to another sequence means that, when aligned, that percentage of bases are the same in comparing the two sequences.

Homology, as described herein, means that the polypeptide sequences that are encoded by the nucleic acids demonstrate a certain relatedness (i.e., there exists regions of conserved amino acids), but not the same amino acid identity. There is complete or 100% homology at a particular amino acid residue when the amino acids of sequences being compared are the same (there is identity) or represent a conservative amino acid substitution (there is homology). A

"conservative amino acid substitution" occurs when a particular amino acid is substituted by an alternate amino acid of similar charge density, hydrophobicity/hydrophilicity, size and/or configuration (e.g., Val for Ile). A "nonconservative amino acid substitution" occurs when a particular amino acid is substituted by an alternative amino acid of differing properties, that is, charge density, hydrophobicity/hydrophilicity, size and/or configuration (e.g., Val for Tyr). The nucleic acid sequences within the scope of the present invention include those nucleic acids which differ in exact sequence from those listed in SEQ ID NO:1 through SEQ ID NO:73 and SEQ ID NO:111 through SEQ ID NO:152 but which encode identical or homologous polypeptide amino acid sequences.

A "primer" is a short polynucleotide, generally with a free 3' -OH group, that binds to a target potentially present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a polynucleotide complementary to the target.

An "adaptor" is a short, partially-duplexed polynucleotide that has a blunt, double-stranded end and a protruding, single-stranded end. It can be ligated, through its double-stranded end, to the double-stranded end of another polynucleotide. This provides known sequences at the ends of thus modified polynucleotides. Often adaptors contain specific sequences for primer binding and/or restriction endonuclease digestion.

A "probe" when used in the context of polynucleotide manipulation refers to a polynucleotide which is provided as a reagent to detect a target potentially present in a sample of interest by hybridizing with the target. Usually, a probe will comprise a label or a means by which a label can be attached, either before or subsequent to the hybridization reaction. Suitable labels include, but are not limited to radioisotopes, fluorochromes, chemiluminescent compounds, dyes, and enzymes.

"Transformation" or "transfection" refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, lipofection, transduction, infection or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host cell genome.

A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the polypeptide, a homologous polypeptide or a fragment thereof. For purposes of this invention, and to avoid cumbersome referrals to complementary strands, the anti-sense (or complementary) strand of such a polynucleotide is also said to encode the

sequence; that is, a polynucleotide sequence that "encodes" a polypeptide includes both the conventional coding strand and the complementary sequence (or strand).

The terms "polypeptide", "oligopeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, it may be interrupted by non-amino acids, and it may be assembled into a complex of more than one polypeptide chain. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component. Also included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), as well as other modifications known in the art.

In the context of polypeptides, a "linear sequence" or a "sequence" is an order of amino acids in a polypeptide in an N-terminal to C-terminal direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polypeptide. A "partial sequence" is a linear sequence of part of a polypeptide which is known to comprise additional residues in one or both directions.

"Recombinant," as applied to a polynucleotide or gene, means that the polynucleotide is the product of various combinations of cloning, restriction and/or ligation steps, and other procedures that result in a construct that is distinct from a polynucleotide found in nature.

A "vector" is a self-replicating nucleic acid molecule that can be used to transfer an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of a nucleic acid molecule into a cell, vectors that function primarily for the amplification of nucleic acid, and expression vectors that function for transcription and/or translation of the DNA or RNA. Also included are vectors that provide more than one of the above functions.

"Expression vectors" are defined as polynucleotides which, when introduced into an appropriate host cell, can be transcribed into a mRNA capable of being translated into a polypeptide(s). An expression vector also comprises control elements operatively linked to the coding region to enable and/or facilitate expression of the polypeptide in the target cell. These can include transcriptional, translational, posttranscriptional, and posttranlational control elements, as are known in the art. An "expression system" usually connotes a suitable host cell comprised of an expression vector that can function to yield a desired expression product.

A “host cell” includes an individual cell or cell culture which can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in genomic or total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected in vivo with a polynucleotide(s) of this invention.

A “cell line” or “cell culture” denotes eukaryotic cells, derived from higher, multicellular organisms, grown or maintained in vitro. It is understood that the descendants of a cell may not be completely identical (either morphologically, genotypically, or phenotypically) to the parent cell. Cells described as “uncultured” are obtained directly from a living organism, and are generally maintained for a limited amount of time away from the organism (i.e., not long enough or under conditions for the cells to undergo substantial replication).

As used herein, “expression” includes transcription and/or translation.

“Heterologous” means derived from (i.e., obtained from) a genotypically distinct entity from the rest of the entity to which it is being compared. For example, a polynucleotide may be placed by genetic engineering techniques into a plasmid or vector derived from a different source, thus becoming a heterologous polynucleotide. A promoter which is linked to a coding sequence with which it is not naturally linked is a heterologous promoter.

An “isolated” or "purified" polynucleotide, polypeptide or cell is one that is substantially free of the materials with which it is associated in nature. By substantially free is meant at least 50%, preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, even more preferably at least 99%, and even more preferably at least 99.9% free of the materials with which it is associated in nature. As used herein, an “isolated” polynucleotide or polypeptide also refers to recombinant polynucleotides or polypeptides, which, by virtue of origin or manipulation: (1) are not associated with all or a portion of a polynucleotide or polypeptide with which they are associated in nature, (2) are linked to a polynucleotide or polypeptide other than that to which they are linked in nature, or (3) do not occur in nature, or (4) in the case of polypeptides, arise from expression of recombinant polynucleotides. Thus, for example, an isolated substance may be prepared by using a purification technique to enrich it from a source mixture. Enrichment can be measured on an absolute basis, such as weight per volume of solution, by specific activity or it can be measured in relation to a second, potentially interfering substance present in the source mixture. Increasing enrichments of the embodiments of this invention are increasingly more preferred.

Thus, for example, a 2-fold enrichment is preferred, 10-fold enrichment is more preferred, 100-fold enrichment is more preferred, 1000-fold enrichment is even more preferred. A substance can also be provided in an isolated state by processes such as chemical synthesis or recombinant expression.

A "reagent" polynucleotide, polypeptide, or antibody, is a substance provided for a reaction, the substance having some known and desirable function in the reaction. A reaction mixture may also contain a "target", such as a polynucleotide, antibody, polypeptide, or assembly of polypeptides that the reagent is capable of reacting with. For example, in some types of diagnostic tests, the presence and/or amount of the target in a sample is determined by adding a reagent, allowing the reagent and target to react, and measuring the amount of reaction product (if any).

"Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of an amplification reaction such as PCR, or the enzymatic cleavage of a polynucleotide by a ribozyme.

When hybridization occurs in an antiparallel configuration between two single-stranded polynucleotides, those polynucleotides are described as "complementary". A double-stranded polynucleotide can be "complementary" to another polynucleotide if hybridization can occur between one of the strands of the first polynucleotide and the second. The degree to which one polynucleotide is complementary with another is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonds with each other, according to generally accepted base-pairing rules of A-T, A-U and G-C.

A "stable duplex" of polynucleotides, or a "stable complex" formed between any two or more components in a biochemical reaction, refers to a duplex or complex that is sufficiently long-lasting to persist between formation of the duplex or complex and subsequent detection, including any optional washing steps or other manipulation that may take place in the interim.

A substance is said to be "selective" or "specific" if it reacts or associates more frequently, more rapidly, with greater duration and/or with greater affinity with a particular cell or substance than it does with alternative cells or substances. An odorant ligand "specifically

binds" to a target if it binds with greater affinity, avidity, more readily, and/or with greater duration than it binds to other substances.

As used herein, "naturally occurring," "native," or "wild type" refers to endogenous polynucleotides and the protein(s) expressed thereby. These terms include full-length and processed polynucleotides and polypeptides. Processing can occur in one or more steps, and these terms encompass all stages of processing. For instance, polypeptides having or lacking a signal sequence are encompassed by the invention. "Non-naturally occurring", "non-native", or "non-wild type" refer to all other polynucleotides and polypeptides.

A "polymerase chain reaction" ("PCR") is a reaction in which replicate copies are made of a target polynucleotide using one or more primers, and a catalyst of polymerization, such as a reverse transcriptase or a DNA polymerase, and particularly a thermally stable polymerase enzyme. Methods for PCR are taught in U.S. Patent Nos. 4,683,195 (Mullis) and 4,683,202 (Mullis et al.). All processes of producing replicate copies of the same polynucleotide, such as PCR or gene cloning, are collectively referred to herein as "amplification."

According to this invention, a "genomic DNA library" is a clone library which contains representative nucleotide sequences from the DNA of a given genome. It is constructed using various techniques that are well known in the art, for instance, by enzymatically or mechanically fragmenting the DNA from an organism, organ, or tissue of interest, linking the fragments to a suitable vector, and introducing the vector into appropriate cells so as to establish the genomic library. A genomic library contains both transcribed DNA fragments as well as nontranscribed DNA fragments.

In comparison, a "cDNA library" is a clone library that differs from a genomic library in that it contains only transcribed DNA sequences and no nontranscribed DNA sequences. It is established using techniques that are well known in the art, i.e., selection of mRNA (e.g. by polyA) making single stranded DNA from a population of cytoplasmic mRNA molecules using the enzyme RNA-dependent DNA polymerase (i.e., reverse transcriptase), converting the single-stranded DNA into double-stranded DNA, cloning the resultant molecules into a vector, and introducing the vector into appropriate cells so as to establish the cDNA library. Alternately, a cDNA library need not be cloned into a vector and/or established in cells, but can be screened using PCR with gene-specific primers, as is well known in the art.

An "individual" is a vertebrate, preferably a mammal, more preferably a human.

General Techniques

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology and biochemistry, which are within the skill of the art. Such techniques are explained fully in the literature, such as: "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture" (R.I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Gene Transfer Vectors for Mammalian Cells" (J.M. Miller & M.P. Calos, eds., 1987); "Current Protocols in Molecular Biology" (F.M. Ausubel et al., eds., 1987 and annual updates); "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994); "Current Protocols in Immunology" (J.E. Coligan et al., eds., 1991).

Basis for identification and description of the polynucleotides and polypeptides

The polynucleotide sequences were identified using oligonucleotide primers which were complementary to OR membrane-spanning regions. A number of different primers were used to elicit a variety of nucleotide sequences which encode polypeptides involved in olfactory sensation. The identification and isolation of nucleotide sequences which encode polypeptides involved in olfactory sensation and the polypeptides that they encode is vital for determining the response of receptors to odorant molecules, the elucidation of scent representations, profiles, or fingerprints, the reproduction of scent representations, profiles, or fingerprints and the editing of scent representations, profiles, or fingerprints.

Polynucleotides encoding polypeptides involved in olfactory sensation

The present invention provides isolated polynucleotides encoding polypeptides which are involved in olfactory sensation, vectors containing these polynucleotides, host cells containing these polynucleotides, and compositions comprising these polynucleotides. These polynucleotides are isolated and/or produced by chemical and/or recombinant methods, or a combination of these methods. The present invention includes polynucleotides isolated from the human olfactory epithelium which encode polypeptides which are involved in olfactory sensation, vectors containing these polynucleotides, host cells containing these polynucleotides, and compositions comprising these polynucleotides. Unless specifically stated otherwise,

"polynucleotides" shall include all embodiments of the polynucleotides of this invention. These polynucleotides are useful as probes, primers, in expression systems, and, in a preferred embodiment, in screening methods as described herein. In one embodiment the polynucleotides of the present invention can be isolated by creating a cDNA library using template RNA from human olfactory epithelium tissue. A detailed example is related in Example 1, below.

The advantage of constructing a cDNA library for isolation of the desired nucleotide sequences is that the likelihood of obtaining pseudogenes is greatly reduced compared to using a genomic DNA library for the same purpose. cDNA libraries contain only mRNA expressed in the tissue used for the construction of the library, in this case, the human olfactory epithelium. The preferred olfactory epithelium tissue should express only those nucleotide sequences which are relevant for olfactory function, thereby excluding nonfunctioning pseudogenes and also GPCRs which may be similar in primary structure (amino acid sequence) but are not encoded in OSNs. As the number of GPCRs utilized in human signal transduction pathways is extremely wide and varied, cDNA libraries constructed using olfactory tissue are preferable for isolating nucleotide sequences that encode polypeptides which are involved in olfactory sensation, inasmuch as genomic libraries can contain abundant nucleotide sequences which encode for a variety of GPCRs performing numerous functions, and are likely to contain pseudogenes.

The isolation of polynucleotide sequences which encode polypeptides involved in olfactory sensation is described in Example 1. Accordingly, this invention provides isolated polynucleotides that contain sequences encoding polypeptides or portions thereof which are involved in olfactory sensation, wherein the polypeptide is at least 10 amino acids in length, and wherein the polynucleotide sequences are depicted in SEQ ID NOs:1-73 and SEQ ID NOs:111-152.

The invention includes modifications to said polynucleotides described above such as deletions, substitutions, additions, or changes in the nature of any nucleic acid moieties. A "modification" is any difference in nucleotide sequence as compared to a polynucleotide shown herein to encode a polypeptide involved in olfactory sensation, and/or any difference in the nucleic acid moieties of the polynucleotide(s), wherein such a modified polynucleotide encodes a polypeptide involved in olfactory sensation or a variant of said polypeptide that is useful in the practice of the invention. Such changes can be useful to facilitate cloning and modify expression of polynucleotides encoding polypeptides which are involved in olfactory

sensation. Such changes also can be useful for conferring desirable properties to the polynucleotide(s), such as stability. The definition of polynucleotide provided herein gives examples of these modifications. Hence, the invention also includes variants of the nucleic acid sequences disclosed herein, which include nucleic acid substitutions, additions, and/or deletions.

The invention also encompasses polynucleotides encoding polypeptides involved in olfactory sensation, including polynucleotides that are full-length, processed, coding, non-coding (including flanking region) or portions thereof, provided that these polynucleotides contain a region encoding at least a portion of a polypeptide involved in olfactory sensation. (That is, the region encodes a functional fragment of an olfactory receptor or other polypeptide involved in olfactory sensation.) Also embodied are the mRNA, cDNA and genomic DNA sequences and fragments thereof that include a polynucleotide sequence comprising a coding sequence for a portion of a polypeptide involved in olfactory sensation.

Genes encoding human olfactory receptors, and optionally including related genomic sequences such as regulatory sequences, can be obtained using olfactory receptor cDNAs as hybridization probes. Under high stringency hybridization conditions, an OR cDNA will hybridize to its cognate OR gene. Use of lower stringency hybridization conditions allows the isolation of OR genes that are related to, but not identical with, the gene corresponding to a particular OR cDNA.

Conditions for hybridization are well-known to those of skill in the art and can be varied within relatively wide limits. Hybridization stringency refers to the degree to which hybridization conditions disfavor the formation of hybrids containing mismatched nucleotides, thereby promoting the formation of perfectly matched hybrids or hybrids containing fewer mismatches; with higher stringency correlated with a lower tolerance for mismatched hybrids. Factors that affect the stringency of hybridization include, but are not limited to, temperature, pH, ionic strength, and concentration of organic solvents such as formamide and dimethylsulfoxide. As is well known to those of skill in the art, hybridization stringency is increased by higher temperatures and/or lower ionic strengths. See, for example, Ausubel et al., supra; Sambrook et al., supra; M.A. Innis et al. (eds.) PCR Protocols, Academic Press, San Diego, 1990; B.D. Hames et al. (eds.) Nucleic Acid Hybridisation: A Practical Approach, IRL Press, Oxford, 1985; and van Ness et al., (1991) Nucleic Acids Res. 19:5143-5151. The degree of stringency can be adjusted not only during a hybridization reaction, but also in post-hybridization washes, as is known to those of skill in the art.

The invention also encompasses polynucleotides encoding polypeptides involved in olfactory sensation, functionally equivalent variants and derivatives of full-length polypeptides involved in olfactory sensation and functionally equivalent fragments. For instance, changes in a DNA sequence that do not change the encoded amino acid sequence, as well as those that result in conservative substitutions of amino acid residues, non-deleterious non-conservative substitutions, one or a few amino acid deletions or additions, and substitution of amino acid residues by amino acid analogs, will not significantly affect properties of the encoded polypeptide. Polypeptides homologous to the polypeptides encoded by the polynucleotides described herein can also be identified using algorithms and methods well-known to those of skill in the art, such as those described in Ausubel, "Current Protocols in Molecular Biology," Chapter 19; see also Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410; Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3:266-272; Madden, T.L., Tatusov, R.L. & Zhang, J. (1996) "Applications of network BLAST server" Meth. Enzymol. 266:131-141; Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402; and Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649-656. A preferred method of determining homology is the BLAST set of similarity search programs (Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410. Polypeptides which are 40% homologous, 50% homologous, 60% homologous, 70% homologous, 80% homologous, 90% homologous, 95% homologous, or 99% homologous to the polypeptides encoded by the polynucleotides described herein are encompassed by the invention.

Nucleotide substitutions that do not alter the amino acid residues encoded can be useful for optimizing gene expression in different systems. Suitable substitutions are known to those of skill in the art and are made, for instance, to reflect preferred codon usage in the particular expression systems. In another example, alternatively spliced polynucleotides can give rise to different functionally equivalent fragments or variants of an polypeptide involved in olfactory sensation. Alternatively processed polynucleotide sequence variants are defined as polynucleotide sequences corresponding to mRNAs that differ in sequence from one another but are derived from the same genomic region, for example, mRNAs that result from: 1) the

use of alternative promoters; 2) the use of alternative polyadenylation sites; and/or 3) the use of alternative splice sites.

Preparation of polynucleotides involved in olfactory sensation

The polynucleotides of this invention can be obtained using chemical synthesis, recombinant methods, or PCR.

Methods of chemical polynucleotide synthesis are well known in the art and need not be described in detail herein. One of skill in the art can use the sequences provided herein and a commercial DNA synthesizer to produce a desired DNA sequence.

For preparing polynucleotides which encode polypeptides involved in olfactory sensation using recombinant methods, a polynucleotide comprising a desired sequence can be inserted into a suitable vector, and the vector in turn can be introduced into a suitable host cell for replication and amplification. Polynucleotides may be inserted into host cells by any means known in the art. Cells are transformed by introducing an exogenous polynucleotide by direct uptake, endocytosis, transfection, F-mating, particle bombardment, liposome mediation, or electroporation. Once introduced, an exogenous polynucleotide can be maintained within the cell as a non-integrated vector (such as a plasmid) or integrated into the host cell genome. The polynucleotide encoding a polypeptide involved in olfactory sensation can be isolated from the host cell by methods well known within the art. See, e.g., Sambrook et al. (1989).

Alternatively, PCR allows amplification of DNA sequences. PCR technology is well known in the art and is described in U.S. Pat. Nos. 4,683,195, 4,800,159, 4,754,065 and 4,683,202, as well as *PCR: The Polymerase Chain Reaction*, Mullis et al. eds., Birkhausw Press, Boston (1994).

RNA can be obtained in a number of ways in an appropriate vector and the vector is transformed into a suitable host cell. When the inserted DNA is transcribed into RNA, the RNA can then be isolated using methods well known to those of skill in the art, as set forth in Sambrook et al., (1989), for example. RNA can also be obtained through in vitro reactions. For example, the polynucleotide, which encodes a polypeptide involved in olfactory sensation, can be inserted into a vector that contains appropriate transcription promoter sequences. Commercially available RNA polymerases will specifically initiate transcription at their promoter sites and continue the transcription process through the adjoining DNA polynucleotides. Placing the polynucleotide sequences which encode polypeptides involved in

olfactory sensation between two such promoters allows the generation of sense or antisense strands of desired RNA.

Cloning and expression vectors comprising polynucleotide sequences encoding polypeptides involved in olfactory sensation

The present invention further includes a variety of vectors containing polynucleotides encoding polypeptides involved in olfactory sensation. These vectors can be used for expression of recombinant polypeptides as well as a source of polynucleotides which encode polypeptides involved in olfactory sensation. Cloning vectors can be used to obtain replicate copies of the polynucleotides, which encode polypeptides involved in olfactory sensation, they contain, or as a means of storing the polynucleotides in a depository for future recovery. Expression vectors (and host cells containing these expression vectors) can be used to obtain polypeptides produced from the polynucleotides they contain. Suitable cloning and expression vectors include any known in the art, e.g., those for use in in vitro, bacterial, mammalian, yeast and insect expression systems. Specific vectors and suitable host cells are known in the art and need not be described in detail herein. For example, see Gacesa and Ramji, *Vectors*, John Wiley & Sons (1994).

Cloning and expression vectors typically contain a selectable marker (for example, a gene encoding a protein necessary for the survival or growth of a host cell transformed with the vector), although such a marker gene can be carried on another polynucleotide sequence co-introduced into the host cell. Only those host cells into which a selectable marker has been introduced will survive and/or grow under selective conditions. Typical selectable markers encode protein(s) that (a) confer resistance to antibiotics or other toxins substances, e.g., ampicillin, neomycin, methotrexate, etc.; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media. The choice of the proper marker gene will depend on the host cell, and appropriate genes for different hosts are known in the art. Cloning and expression vectors also typically contain a replication system recognized by the host.

Suitable cloning vectors may be constructed according to standard techniques, or may be selected from a large number of cloning vectors available in the art. While the cloning vector selected may vary according to the host cell intended to be used, useful cloning vectors will generally have the ability to self-replicate in an appropriate host, may possess a single target for one or more particular restriction endonucleases, and/or may carry genes for a marker

that can be used in selecting clones containing the vector. Suitable examples include plasmids and bacterial viruses, e.g., pUC18, pUC19, m13mp18, m13mp19, pBR322, pMB9, ColE1, pCR1, RP4, phage DNAs, and shuttle vectors such as pSA3 and pAT28. These and many other cloning vectors are available from commercial vendors such as BioRad, Stratagene, and Invitrogen.

Expression vectors generally are replicatable polynucleotide constructs that contain a polynucleotide encoding an polypeptide involved in olfactory sensation of interest. The polynucleotide, which encodes a polypeptide involved in olfactory sensation, encoding the polypeptide is operatively linked to suitable transcriptional controlling elements, such as promoters, enhancers and terminators. For expression (i.e., translation), one or more translational controlling elements are also usually required, such as ribosome binding sites, translation initiation sites, and stop codons. These controlling elements (transcriptional and translational) may be derived from the gene encoding polypeptides involved in olfactory sensation, or they may be heterologous (i.e., derived from other genes and/or other organisms). A polynucleotide sequence encoding a signal peptide can also be included to allow a polypeptide involved in olfactory sensation to cross and/or lodge in cell membranes or be secreted from the cell. A number of expression vectors suitable for expression in eukaryotic cells including yeast, insect, avian, plant and mammalian cells are known in the art. Common vectors, such as YEp13 and the Sikorski series pRS303-306, 313-316, 423-426 can also be used. Vectors pDBV52 and pDBV53 are suitable for expression. Another example of an expression vector/host cell system is the baculovirus (*e.g.*, nuclear polyhedrosis virus)/insect cell (*e.g.*, sf9 cells) system.

Human olfactory receptor polypeptides are expressed from olfactory receptor cDNA by methods well-known to those of skill in the art. A cDNA or portion thereof is inserted in an expression vector using standard molecular cloning techniques. Coupled in vitro transcription and translation of such a vector results in expression of the OR protein encoded by the cDNA. In vivo expression of a OR polypeptide is accomplished by inserting an OR cDNA into a eucaryotic or procaryotic expression vector, of which many are known in the art, to genereate an OR expression construct. The OR expression construct is introduced into an appropriate host cell in which the OR sequences are expressed (by transcription and translation) and optionally secreted, and the expressed OR polypeptide is obtained from the cell growth medium and/or from cell lysates.

A number of expression vectors are known in the art. Prokaryotic expression vectors include, but are not limited to, T7 RNA polymerase/T7 promoter-based vectors, bacteriophage λ-based vectors and various types of fusion vectors. Fusion vectors include, but are not limited to, lacZ and trpE fusion vectors, maltose binding protein fusion vectors, glutathione-S-transferase fusion vectors, and thioredoxin fusion vectors. Baculovirus-based vectors are used for expression in insect cell systems. Expression in mammalian cells (such as HEK, COS and CHO cells) utilizes vectors containing a mammalian origin of replication (such as, for example, a SV40 origin), an efficient promoter (optionally including one or more enhancer sequences), mRNA processing signals (e.g., splice sites and polyadenylation sites), one or more selectable markers, and optionally a prokaryotic replicon to allow propagation and manipulation of the construct in prokaryotic cells. Alternatively, expression in mammalian cells is achieved through the use of any of a number of mammalian viral vectors including, but not limited to, retroviruses, lentiviruses, Semliki Forest viruses, vaccinia viruses, adenoviruses and adeno-associated viruses.

Vectors containing the polynucleotides of interest can be introduced into the host cell by any of a number of appropriate means, including electroporation, direct injection, transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; and infection (where the vector is an infectious agent, such as a virus). The choice of means of introducing vectors or polynucleotides encoding polypeptides involved in olfactory sensation will often depend on the host cell, as will be well known to those of skill in the art.

Host cells transformed with polynucleotides encoding polypeptides involved in olfactory sensation

Another embodiment of this invention are host cells transformed with (i.e., comprising) polynucleotides encoding polypeptides involved in olfactory sensation, and/or vectors having polynucleotide(s) sequences encoding polypeptides involved in olfactory sensation, as described above. Both prokaryotic and eukaryotic host cells may be used. Prokaryotic hosts include bacterial cells, for example *E. coli*, *B. subtilis*, and mycobacteria. Among eukaryotic hosts are yeast, insect, avian, plant and mammalian cells. Host systems are known in the art and need not be described in detail herein.

The host cells of this invention can be used, *inter alia*, as repositories of polynucleotides encoding polypeptides involved in olfactory sensation, and/or vehicles for

production of polynucleotides encoding polypeptides involved in olfactory sensation, and/or polypeptides involved in olfactory sensation . They may also be used as vehicles for *in vivo* delivery of polypeptides involved in olfactory sensation .

Uses for and methods using polynucleotides encoding polypeptides involved in olfactory sensation

To determine whether a vector containing polynucleotides is capable of expressing in eukaryotic cells, cells such as, for example, COS-7 (primate origin), CHO (rodent origin), HEK-293 (human origin), or HeLa (human origin) cells can be transfected with the vector. Expression of a polypeptide(s) encoded by the vector is then determined by, for example, RIA, ELISA, immunofluorescence of fixed cells, or western blotting of cell lysate using an antibody as a probe. Antibodies can be obtained using, as immunogen, peptide sequences synthesized from the protein sequences encoded by the known polynucleotide sequence. Polypeptides can be purified by, for example, phase partitioning, affinity methods, gel filtration and ion exchange, as well as additional methods known by those skilled in the art. Further characterization of the expressed polypeptide can be achieved by purification of the polypeptide using techniques known in the art.

Polypeptides involved in olfactory sensation

The present invention encompasses polypeptides involved in olfactory sensation. Expression of said polypeptides is localized in the olfactory neurons located in the olfactory epithelium, as described earlier. The polypeptides may comprise any novel sequence encoded by a nucleotide sequence as depicted in SEQ ID NO:1 through SEQ ID NO:73 and SEQ ID NO:111 through SEQ ID NO:152.

The invention includes modifications to polypeptides involved in olfactory sensation including functionally equivalent fragments of the polypeptides involved in olfactory sensation which do not significantly affect their properties and variants which may have enhanced or decreased activity. Collectively, these modifications may be termed "analogs" of or a fragment of polypeptides involved in olfactory sensation. Modification of polypeptides is routine practice in the art and need not be described in detail herein. Examples of modified polypeptides include polypeptides with conservative substitutions of amino acid residues, one or more deletions or additions of amino acids which do not significantly deleteriously change the functional activity, or use of chemical analogs. Amino acid residues which can be conservatively substituted for

one another include but are not limited to: glycine/alanine; valine/isoleucine/leucine; asparagine/glutamine; aspartic acid/glutamic acid; serine/threonine; lysine/arginine; and phenylalanine/tyrosine. Such conservative substitutions are known in the art, and preferably, the amino acid substitutions would be such that the substituted amino acid would possess similar chemical properties as that of the original amino acid. These polypeptides also include glycosylated and non-glycosylated polypeptides, as well as polypeptides with other post-translational modifications, such as, for example, glycosylation with different sugars, acetylation, and phosphorylation. Amino acid modifications can range from changing or modifying one or more amino acids to complete redesign of a region. Other methods of modification include using coupling techniques known in the art, including, but not limited to, enzymatic means, oxidative substitution and chelation. Modified polypeptides involved in olfactory sensation are made using established procedures in the art.

The invention also encompasses fusion proteins comprising one or more polypeptides involved in olfactory sensation. For purposes of this invention, an fusion protein contains one or more polypeptides involved in olfactory sensation and another amino acid sequence to which it is not attached in the native molecule, for example, a heterologous sequence or a homologous sequence from another region. Useful heterologous sequences include, but are not limited to, sequences that provide for secretion from a host cell, intracellular trafficking, and stability/degradation. Other useful heterologous sequences are ones which facilitate purification. Examples of such sequences are known in the art and include those encoding epitopes such as Myc, HA (derived from influenza virus hemagglutinin), His-6, or FLAG. Other heterologous sequences that facilitate purification are derived from proteins such as glutathione S-transferase (GST), maltose-binding protein (MBP), or the Fc portion of immunoglobulin.

Preparation of polypeptides involved in olfactory sensation

The polypeptides of this invention can be made by procedures known in the art. The polypeptides can be produced by recombinant methods (i.e., single or fusion polypeptides) or by chemical synthesis. Polypeptides, especially shorter polypeptides up to about 50 amino acids, are conveniently made by chemical synthesis. Methods of chemical synthesis are known in the art and are commercially available. For example, a polypeptide can be produced by an automated polypeptide synthesizer employing the solid phase method. Polypeptides can also be made by chemical synthesis using techniques known in the art.

Polypeptides can also be made by expression systems, using recombinant methods. The availability of polynucleotides encoding polypeptides permits the construction of expression vectors encoding intact (i.e., native) polypeptide, functional equivalents and functional fragments thereof, modified forms or recombinant forms. A polynucleotide encoding the desired polypeptide, or a fusion protein, can be ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems can be used. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification or isolation of the polypeptides expressed in host systems can be accomplished by any method known in the art (e.g. partitioning exclusion, ion exchange chromatograph, gel filtration, etc.). Other controlling transcription or translation segments, such as signal sequences that direct the polypeptide to a specific cell compartment (i.e., for secretion), can also be used. Examples of prokaryotic host cells are known in the art and include, for example, *E. coli* and *B. subtilis*. Examples of eukaryotic host cells are known in the art and include yeast, avian, insect, plant, and animal cells such as COS7, HeLa, CHO, HEK-293 and other mammalian cells.

Alternatively, in vitro expression systems may also be used to produce polypeptides involved in olfactory sensation. A plasmid containing a polynucleotide encoding polypeptides involved in olfactory sensation, under the control of an appropriate promoter, can be transcribed and the resultant RNA translated in vitro through the use of commercially available reagents. Such methods can be used to produce relatively pure samples of the polypeptide and are known in the art.

Preferably, the polypeptides are at least partially purified from other cellular constituents. In one embodiment, the polypeptides are at least 70%, more preferably at least 80%, even more preferably at least 90% or most preferably at least 95% pure. In this context, purity can be calculated as a weight percent of the total protein content of the preparation. More highly purified polypeptides may also be obtained and are encompassed by the present invention. Methods of protein purification are known in the art and are not described in detail herein. For membrane-bound proteins, the lipid content of the preparation, which is required to maintain the structure and function of the protein, is excluded from the purity calculation. That is, if a preparation weighing 10 mg has 5 mg lipid, 4 mg of desired protein, and 1 mg of undesired proteins, the purity is calculated as 80% (desired protein content divided by total protein content). Preparations of biological or synthetic molecules suitable for maintaining structure and function of membrane proteins are described in Etemadi AH (1985) *Adv Lipid*

Res 1985;21:281-428; Villalobo A (1990) *Biochimica Et Biophysica Acta*, 1017(1):1-48; Montal M (1987) *Journal Of Membrane Biology* 98(2): 101-115; Scotto AW et al. (1987) *Biochemistry* 26(3): 833-839; Jain MK and Zakim D (1987) *Biochimica Et Biophysica Acta* 906(1): 33-68; Czerski L and Sanders CR (2000) *Anal Biochem* 284(2):327-33 (lipid-detergent mixtures or "bicelles"); Hrafnsdottir S and Menon AK (2000) *J Bacteriol* 182(15):4198-206 (proteoliposomes); Puu G et al. (2000) *Biosens Bioelectron* 15(1-2):31-41 (protein-lipid preparations on solid surfaces); Schafmeister CE et al. (1993) *Science* 262(5134):734-8 ("peptitergents").

Uses of polypeptides involved in olfactory sensation

The polypeptides of this invention have a variety of uses. They can be used, for example, to screen odorant ligands in order to determine the scent representations, scent profiles or scent fingerprints of particular odorant molecules and further to characterize the effect of functional groups and chemical characteristics on perceived smell. Methods for screening odorant compounds using odorant receptors in neuronal cells are known in the art (Firestein et al., WO 98/50081; Duchamp-Viret *et al., Science* 1999, **284** 2171-2174; Sato *et al., J. Neurophys.* 1994 **72** 2980-2989; Malnic *et al, Cell* 1999 **96** 713-723; Zhao *et al., Science* 1998 **279**, 237-242). There are also methods which can be employed to screen odorant compounds which do not require neuronal cells and are known in the art (Kauvar et al., U. S. Pat. No. 5,798,275; Kiefer *et al., Biochemistry* 1996 **35** 16077-16084; Krautwurst *et al., Cell* 1998 **95** 917-926),

Analysis of the scent can be performed in a number of ways. Various embodiments of the scent analysis system are presented. Examples of how these embodiments might operate are also presented, although it should be emphasized that the invention is not limited by any particular theory of olfactory perception or scent analysis.

<u>*Olfactory Space*</u>

The sensory subsystem comprises a series of olfactory receptors, which selectively bind with the chemical component(s) making up the scent. The scent can be characterized in terms of which of the approximately 1,000 olfactory receptors the scent component(s) bind to, and the strength of the interaction of the component(s) with those receptors. Each olfactory receptor can be considered an orthogonal basis vector; the entire set of olfactory receptors can be considered a set of basis vectors spanning "olfactory space." This is analogous to vectors

pointing along the x, y, and z directions in three-dimensional space, where any point in space can be represented by a combination of the x, y, and z basis vectors (with each of the x, y, and z vectors multiplied by the appropriate scalar quantity). The intensity of interaction of a scent with an olfactory receptor determines the magnitude of the vector along that particular "axis" in olfactory space. Thus, every scent can be uniquely described by a vector representation in olfactory space.

A representation of a scent in such a manner that the scent can later be re-created is defined as scent profiling. The aforementioned vector representation is one example of a scent profile.

Primary Scents

For the purposes of this invention, a receptor primary scent component is defined as a chemical that interacts with one and only one scent receptor. A receptor complex scent component is defined as a chemical that interacts with more than one scent receptor; the receptor complex scent component can interact with each of the scent receptors to different degrees, to equal degrees, or can interact with some receptors to the same degree and others to different degrees.

Olfactory receptors are proteins which fall in the class of seven transmembrane domain G protein-coupled receptors, and are found in olfactory neurons *in vivo*. Binding of an odorant to an olfactory receptor causes second messenger systems to become activated or inhibited in the cell, leading to increased cellular production of second messenger molecules such as cyclic AMP. These second messenger systems in turn lead to the depolarization of the olfactory neuron, or other changes in the state of the neuron, which provides the signal to the nervous system that the odorant has been detected.

With a complete set of receptor primary scent components, any scent can be re-created with the knowledge to the degree to which it interacts with each olfactory receptor. The instant invention encompasses such complete sets of receptor primary scent components. Other embodiments of the invention encompass sets of receptor primary scent component chemicals which provide the ability to re-create a particularly desired subset of scents, but not necessarily all possible scents. Still more embodiments encompass sets of receptor primary scent component chemicals which provide the ability to approximate particular scents, while not necessarily exactly re-creating the interaction profile of the particular scents.

In some cases, a receptor complex scent will be an acceptable approximation to a receptor primary scent. That is, if a given receptor complex scent interacts with a first scent receptor strongly, but interacts with other scent receptors less strongly, it can be considered an approximation to a receptor primary scent component for the first receptor. Such a receptor complex scent component is described by the term receptor quasi-primary scent component. One embodiment of the invention encompasses sets of receptor quasi-primary scent component chemicals suitable for re-creating all scents. Another embodiment of the invention encompasses sets of receptor quasi-primary scent component chemicals suitable for re-creating a particularly desired subset of scents, but not necessarily all possible scents. Yet another embodiment encompasses sets of receptor quasi-primary scent component chemicals which provide the ability to approximate particular scents, while not necessarily exactly re-creating the interaction profile of the particular scents.

The identification of receptor primary or quasi-primary scent component chemicals provides the most conceptually straightforward method of re-creating scents. However, another embodiment of the invention encompasses the use of receptor complex scent components for re-creating scents. An example of such an embodiment would be re-creation of a scent that activates olfactory receptors designated OR1, OR2, OR3, OR4, OR5 and OR6 (for the sake of illustration, it is assumed that the olfactory receptors are stimulated to an equal extent). If one is in possession of two receptor complex scent component chemicals (RCSC's) where RCSC1 activates OR1 and OR5, and RCSC2 activates OR2, OR3, OR4, and OR6, then one can reproduce the original scent by mixing RCSC1 and RCSC2 to re-create the original olfactory receptor activation profile. In practice, the profiles of various receptor complex scent components will be much more complicated than the forgoing example, and components which inhibit olfactory activation as well as stimulate activation can be included in the sets. However, once receptor activation profiles of sufficient receptor complex scent components are known, computer algorithms can be utilized to create the appropriate combination of receptor complex scent components. Using vector representations of the olfactory receptor activation profiles for a set of receptor complex scent components, one can create linear combinations of such receptor complex scent components in order to represent a particular scent. For the example given above, such a vector representation would look like (1, 0, 0, 0, 1, 0) for the first receptor complex scent component and (0, 1, 1, 1, 0, 1) for the second receptor

complex scent component, while the vector representation of the scent to be re-created is (1, 1, 1, 1, 1, 1). If x_1 and x_2 are the relative proportions of the first receptor complex scent component and the second receptor complex scent component, respectively, to be combined to re-create the scent, then the problem can be represented as a series of linear equations:

$$
\begin{array}{ccccc}
1x_1 & + & 0x_2 & = & 1 \\
0x_1 & + & 1x_2 & = & 1 \\
0x_1 & + & 1x_2 & = & 1 \\
0x_1 & + & 1x_2 & = & 1 \\
1x_1 & + & 0x_2 & = & 1 \\
0x_1 & + & 1x_2 & = & 1
\end{array}
$$

and the solutions for x_1 and x_2 are $x_1 = 1$, $x_2 = 1$. Solutions to systems of linear equations have been thoroughly studied and many algorithms are available for implementation on computers, including algorithms which evaluate the accuracy of an approximate solution when an exact solution cannot be determined. (See, e.g., Dettman, J.W., *Introduction to Linear Algebra and Differential Equations*, Dover Pubs., 1986; Press W.H. et al., *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, 1993; Vetterling (ed.) *Numerical Recipes in C: The Art of Scientific Computing/Disk V 2.02*, Cambridge University Press, 1997.) These methods can also be used to determine whether a set of receptor complex scent components is suitable for re-creating a given scent. For example, if the scent to be recreated is represented by the vector (1, 1, 1, 1, 1, 2), there will be no solution to the resulting system of linear equations using the two receptor complex scent components in the illustration above. In this instance, one or more additional receptor scent components will need to be identified in order to be able to re-create the scent in terms of the receptor primary scent components. Alternatively, the scent represented by (1, 1, 1, 1, 1, 1) may be an acceptable approximation to the scent represented by (1, 1, 1, 1, 1, 2). Integers are used in this example for clarity, but the vectors can contain any real number representing a measured intensity; for example, (1.1, 0.997, 1.08, 1.2, 0.88888…, 2.00001) may be an acceptable approximation to the scent represented by (1, 1, 1, 1, 1, 2).

It will be readily appreciated that the choice of a complete set of receptor primary, quasi-primary, or complex scent component chemicals (capable of generating all scents) versus a partial set of receptor primary, quasi-primary, or complex scent component chemicals (capable of generating, exactly or approximately, a subset of scents) depends on the application for which scent re-creation is desired.

A special category of receptor scent components are chemicals which bind to a receptor without activating it. If these non-activating chemicals prevent chemicals which do activate the receptors from binding, the non-activating chemicals act to "turn off" those receptors. These non-activating chemicals, or receptor binding antagonists, are particularly useful in editing scents, as they can be added to a scent to attenuate or eliminate particular aspects of the scent. In the vector example above, if a particular receptor antagonist blocks OR2, OR3, and OR4, but not OR1, OR5 or OR6, then it can be represented in vector format as (0, -1, -1, -1, 0, 0). In the reproduction of (1, 1, 1, 1, 1, 2) from the vectors (1, 0, 0, 0, 1, 0) and (0, 1, 1, 1, 0, 1), the following combination can be used: 1 x (1, 0, 0, 0, 1, 0) + 2 x (0, 1, 1, 1, 0, 1) + 1 x (0, -1, -1, -1, 0, 0) to yield the vector (1, 1, 1, 1, 1, 2). In some instances, enough of a particular receptor binding antagonist is used to eliminate any possibility of activation by a receptor scent component, in which case the vector entry for the receptor(s) which are blocked by that antagonist contains 0 in the vector position corresponding to that receptor(s).

Perceptive primary scents are defined as scents that give a single scent perception, for example, the scent "lemon" as perceived by a human. A perceptive primary scent can be composed of one or more receptor primary scent components, one or more receptor complex scent components, or a mixture of one or more receptor primary scent components and one or more receptor complex scent components. Since perceptive primary scents are to some extent subjective, identification of perceptive primary scents can be performed by using a panel of subjects who evaluate and describe scents. A perceptive complex scent is made up of more than one perceptive primary scent. The boundaries between a perceptive primary scent and a perceptive complex scent are also to some extent subjective; for example, one person may describe a scent as "pizza," while another person may describe the same scent as "sausage, cheese and tomato sauce." That is, one person may perceive a scent as a perceptive primary scent for "pizza," while another person may perceive the same scent as a perceptive complex scent made up of several individual perceptive primary scents. In order to standardize perceptive scents, a panel of five or more, preferably ten or more, more preferably fifty or

more, still more preferably one hundred or more, people can be surveyed to label various perceptive scents. When a plurality, preferably a majority, more preferably 66 2/3 % or greater, still more preferably 95 % or greater, even more preferably 99% or greater, of the panel identifies a scent as the same scent (e.g., of a panel of 100 people, 95 describe a scent as "pizza," while the other 5 describe the scent otherwise), the scent can be labeled as a perceptive scent (the perceptive scent can be primary or complex, depending on whether the panel identifies it as a single scent or a mixture of scents).

In fields where existing classification schemes already exist, the perceptive primary and complex scents can be indexed according to those schemes. For example, the SFP (Société Française des Parfumeurs) has drawn up a classification system based on 5 main groups, sub-divided into classes. Such a classification can be used for selecting perceptive primary scents and used as guides for combining the scents.

Selecting Chemicals for Scent Re-creation

A scent which has been represented as a set of basis vectors in olfactory space can in principle be re-created simply by mixing the receptor primary scent components, receptor quasi-primary scent components, or receptor complex scent components needed to interact the olfactory receptors in the same pattern as the original scent. Such an approach requires 1) a method to generate a representation of the original scent in olfactory space, and 2) suitable receptor primary scent component chemicals which can be mixed in the appropriate manner.

Identification of receptor scent components can be performed by various methods. One such method assays the interaction of candidate components with each olfactory receptor. The receptors can be expressed *in vitro* and assays can be set up to monitor the interaction of various candidate components with each individual receptor. Chemicals which interact with one and only one olfactory receptor are receptor primary scent components, while chemicals which interact with more than one olfactory receptor are receptor complex scent components (and can possibly be receptor quasi-primary scent components, depending on the interaction profile it displays with the olfactory receptors). Such an approach can use methods known in the art, for example those of Breer *et al,* Ann. N. Y. Acad. Sci. (1998) 855:175-81 or Malnic *et al.*, *Cell* (1999) 96(5):713-23. Breer *et al.* expressed olfactory receptors in Sf9 cells and evaluated the second-messenger response to various odorants. Malnic *et al.* isolated olfactory neurons from mice and utilized calcium imaging to study the response of the neurons to different odorants, while using RT-PCR to determine which olfactory receptor was expressed

in the neuron under study. U.S. Patent No. 5,798,275 describes a method for evaluating interaction of compounds with members of a reference panel of proteins. WO 98/50081 discloses methods for detecting particular odorant ligand specificity for particular odorant receptors in nasal epithelium tissue of mammals such as rats and mice.

Selection of Receptor Primary Scents by in silico Methods

An alternative method utilizes *in silico* screening techniques--that is, computer simulation methods--for selecting candidate components. Protein-ligand screening can be used to select compounds which bind to particular receptors in order to identify receptor primary scent components. Examples of such programs are DOCK, AutoDock, GOLD, FlexX, LUDI, GROWMOL, and HOOK. (See Wang, J., Kollman, P.A., Kuntz I.D., "Flexible ligand docking: a multistep strategy approach," *Proteins* 36(1):1-19 (1999) and references therein.) These programs function by taking a protein structure and either matching compounds of known structure to the protein structure to determine the protein-ligand interaction, or by "growing" a molecule in the active site or binding site of a protein to determine what molecule will best interact with the protein.

Olfactory receptor proteins are membrane proteins, and experimental determination of the three-dimensional structures of membrane proteins has lagged the corresponding structural determination of water-soluble proteins for various reasons. However, alternative methods for constructing the three-dimensional structures of proteins are available. The primary (amino acid) sequences of many olfactory receptors are known. This information can be used to model a three-dimensional structure of a receptor protein using various algorithms and computer programs known in the art. The resulting model structure can then be used as the basis for evaluating interaction of candidate components with the receptor.

Alternatively, given known chemical structures which give rise to a particular odor, analysis of the structures can indicate the particular portion of the chemical structure which is responsible for the odor. This is analogous to "pharmacore analysis" used in medicinal chemistry to determine the important portion of drugs.

Methods for developing compounds which bind to receptors and other proteins of known structure, and determining interactions between ligands and receptors, are described in various references The DOCK program evaluates the fit of a ligand into a protein molecule of known structure (see Gschwend, D.A., Good, A.C. and Kuntz, I.D., "Molecular Docking Towards Drug Discovery", *J. Mol. Recognition* 9, 175-86 (1996); Kuntz, I.D., Meng, E.C., and

B.K. Shoichet, "Structure-Based Strategies For Drug Design and Discovery", *Acc. Chem. Res.* 27, 117-123 (1994); and Kuntz, I.D., "Structure-based strategies for drug design and discovery", *Science* 257, 1078-1082 (1992); see also http://www.cmpharm.ucsf.edu/kuntz/dock.html). Using a known (or modeled) structure of an olfactory receptor, DOCK can be used to screen for compounds which bind to the receptor. The program AMBER (see Cornell, WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW and Kollman PA. "A second generation force field for the simulation of proteins and nucleic acids," *Journal of the American Chemical Society* 117, 5179-5197 (1995); Computer Simulation of Biomolecular Systems, A. Wilkinson, P. Weiner, W. Van Gunsteren, eds. Volume 3, p. 83-96, P. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot and A. Pohorille; Bayly CI, Cieplak P, Cornell WD and Kollman PA. "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model," *Journal of Physical Chemistry* 97(40), 10269-10280 (1993); Cornell WD, Cieplak P, Bayly CI and Kollman PA. "Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation," *Journal of the American Chemical Society* 115(21), 9620-9631 (1993); see also http://www.amber.ucsf.edu/amber/amber.html) can be used to calculate more precise interaction energies between candidate ligands. Other examples of such methods are described in, for example, U.S. Patent No. 5,866,343, directed to determining the energetically favorable binding site between two molecules; U.S. Patent No. 5,854,992, a system and method for structure-based drug design which takes into account binding free energy as it "grows" candidate molecules into a receptor binding site; and U.S. Patent No. 5,495,423, which describes a method for ligand design (principally applicable to peptidic ligands).

The foregoing methods typically depend on a known three-dimensional structure for the receptor. When such a structure cannot or has not been determined experimentally, a structure can be modeled using computer algorithms. Blundell TL, Sibanda BL, Sternberg MJ, Thornton JM, "Knowledge-based prediction of protein structures and the design of novel molecules," *Nature* 326(6111):347-52 (1987); Shortle D, "Structure prediction: The state of the art," *Curr Biol* 9(6):R205-9 (1999), Morea V, Leplae R, Tramontano A, "Protein structure prediction and design," *Biotechnol Annu Rev* 4:177-214 (1998) and Onuchic JN, Luthey-Schulten Z, Wolynes PG, "Theory of protein folding: the energy landscape perspective," *Annu Rev Phys Chem* 48:545-600 (1997) address various methods of predicting protein structure from sequence data.

Various implementations for predicting protein structure from amino acid sequences are discussed in U.S. Patent Nos. 5,878,373 and 5,884,230.

If the structure, or even the identity, of the targeted receptor cannot be determined, alternative computational techniques can be used to generate information regarding possible ligands which will interact with the receptor. Quantitative structure-activity relationships (QSAR; see Green, S.M. and Marshall, G.R., "3-D QSAR: A current perspective," Trends Pharmacol Sci 16:285 (1995); and 3D QSAR in Drug Design: Theory, Methods and Applications, Kubinyi, H. Ed.; Escom, Leiden.), including QSAR refinements such as comparative molecular field analysis (ComFA) (Cramer, R. D. et al. "Comparative Molecular Field Analysis ComFA 1. Effect Of Shape On Binding Of Steroids To Carrier Proteins," *J. Am. Chem. Soc.* 110: 5959 (1988)); and pharmacophore mapping (Martin YC, Bures MG, Danaher EA, DeLazzer J, Lico I, Pavlik PA, "A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists," *J Comput Aided Mol Des* 7(1):83-102 (1993)) have been used to design pharmacophores that can interact with the receptor. U.S. Patent No. 5,699,268 provides a method for producing computer-simulated receptors which functionally mimic biological receptors; the simulated receptors are essentially abstractions of structurally useful information from compounds which are known to interact with a receptor. U.S. Patent No. 5,901,069 describes a method of automatically refining a set of chemicals using structure/activity data. U.S. Patent No. 5,862,514 describes a method of simulating synthesis of compounds of desired biological activity and evaluating their activity via further simulations.

Application of structure-function relationships to classification of odors has been described by Chastrette M., Rallet E. "Structure-minty odour relationships: Suggestion of an interaction pattern," *Flavour and Fragrance Journal*, 13(1):5-18 (1998); Chastrette M., De Saint Laumer J.Y.,; Peyraud J.F., "Adapting the structure of a neural network to extract chemical information. Application to structure-odour relationships," *SAR QSAR Environ Res* 1 (2-3):221-231 (1993), Chastrette M., "Trends in structure-odor relationships," *SAR QSAR Environ Res* 6(3-4):215-254 (1997) and Jain et al., "A shape-based machine learning tool for drug design," *J Comput Aided Mol Des* 8(6):635-652 (1994). These methods can be useful in determining the "chemical distance" between odors. For example, isoamyl acetate is typically experienced as a banana-like odor, while octyl acetate is typically experienced as an orange-like odor, which gives a measure of how the chain length of the alkoxy portion of the ester influences perception.

Olfactory Receptors and Libraries of Olfactory Receptors

The olfactory receptors of the invention can be used to analyze and describe the interaction of scent odorant molecules with each receptor. This can be done individually, receptor-by-receptor and odorant molecule by odorant molecule. However, a combinatorial approach provides a much more powerful method of analyzing and describing the interaction of scent odorant molecules with olfactory receptors.

In one embodiment, the invention comprises libraries of olfactory receptors. These libraries are used to screen compositions for interaction with receptors. A composition can be a single compound (essentially a pure chemical), or a mixture of two or more compounds or chemicals. The compositions can be presented to the library in vapor form, or in solutions, typically aqueous solutions.

The method for determining the binding pattern of a composition with olfactory receptors comprises the steps of: exposing the composition to an olfactory receptor library; and determining whether the composition binds to each olfactory receptor of the library, thereby determining the overall binding patter of the composition. While it is desirable to determine whether the composition binds to each of the olfactory receptors, in certain cases, determining the binding pattern to a subset of the receptors is suitable. Such a situation can arise if the complete pattern is not needed, or if the experiment cannot determine binding to a receptor for a particular reason. (Determining the binding to a subset is equivalent to reducing the olfactory receptor library to that subset of receptors.)

Typically, the libraries are prepared as arrays, where the position of each olfactory receptor is known on the array. The arrays can take the form of multiwell plates, solid substrates such as chips or wafers, or any other form allowing identification of the receptor location. The arrays can be prepared in order to simply assess binding, or can be prepared in order to assess degree of activation as described above, using, for example, the technique of Malnic *et al.*, *Cell* 1999 **96**, 713-723. Alternatively, an *in silico* array of structures can be prepared, using the known primary structure of the receptors and the modeling techniques described above.

The libraries contain at least two olfactory receptors. In increasing order of preference, the libraries contain at least 5, 10, 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1500, 1600, 1800, or 2000 olfactory receptors. The

receptors are presented as ordered arrays. For example, a 96-well plate can contain 96 receptor preparations. Upon exposure to a composition, the plate can be scanned, and the response of each receptor in each well can be evaluated. This leads to a 96-element vector description of the composition in terms of those 96 olfactory receptors.

In one embodiment, binding to the olfactory receptors is assessed. In another embodiment, the approximate binding constant of the composition to the olfactory receptors is determined. In yet another embodiment, the degree of activation of the olfactory receptor by the composition is determined. For receptor antagonists, binding will occur, but no activation will occur; the invention embraces the identification of such antagonists.

The compositions for use are varied. A set of all volatile compounds can be used. A standard set of perfumes or odorants can be used. A set of commercially used scents can be used. Sets of compounds particularly useful in the invention are disclosed in co-pending United States Patent Application Serial No. 09/620,753. However, it must be emphasized that the invention is not limited to any one set or classification of compounds.

Preferred subsets of olfactory receptor polynucleotide sequences include:

SEQ ID NOS: 163, 331, 414, 425,672, 762, 919, and 1027;

SEQ ID NOS: 809 and 1067;

SEQ ID NO: 744;

SEQ ID NOS: 207, 336, 441, and 615;

SEQ ID NOS: 157, 168, 197, 221, 250, 334, 340, 412, 413, 459, 491, 618, 690, 694, 759, 760, 761, 767, 819, 860, 872, 873, 917, 936, 939, 940, 947,952, 958, 959, 1023, 1034, 1038, 1043, and 1044;

SEQ ID NOS: 783, 785, 882, 888, 922, and 925;

SEQ ID NOS: 707, 748, 752, 755, 756, 790, and 997;

SEQ ID NOS: 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, and 1084;

SEQ ID NOS: 163, 239, 331, 335, 368, 381, 385, 414, 425, 514, 572, 596, 603, 628, 638, 642, 672,674, 689, 744, 762, 809, 835, 885, 896, 919, 920, 938, 948, 972, 999, 1007, 1014, and 1027;

SEQ ID NOS: 164, 173, 176, 180, 182, 184, 185, 188, 190, 194,207, 210, 213, 214, 215, 217, 219, 220, 223, 226, 227, 229, 230, 234, 235, 240, 249, 255, 265, 270, 273, 274,

276, 277, 279, 281, 289, 291, 293, 294, 298, 302, 307, 311, 318, 319, 321, 330, 336, 339, 341, 342, 343, 348, 351, 356, 359, 361, 365, 366, 367, 368, 370, 372, 373, 374, 375, 376, 378, 379, 380, 382, 383, 384, 385, 388, 391, 392, 393, 398, 400, 401, 403, 408, 420, 423, 427, 428, 431, 434, 435, 438, 439, 440, 441, 447, 448, 450, 455, 458, 464, 465, 468, 471, 473, 474, 475, 478, 479, 481, 482, 484, 485, 492, 494, 499, 502, 508, 511, 512, 513, 515, 526, 532, 534, 541, 543, 545, 546, 550, 552, 553, 557, 558, 560, 563, 564, 568, 572, 576, 582, 583, 584, 585, 586, 588, 599, 600, 605, 606, 607, 608, 609, 610, 615, 620, 621, 631, 632, 636, 638, 640, 642, 645, 648, 650, 651, 652, 654, 656, 657, 661, 662, 664, 668, 679, 680, 686, 687, 689, 691, 696, 699, 700, 702, 706, 713, 720, 721, 723, 729, 734, 738, 745, 768, 772, 773, 775, 791, 798, 799, 823, 857, 898, 900, 901, 903, 914, 931, 933, 937, 941, 945, 948, 956, 965, 969, 983, 992, 993, 994, 999, 1003, 1005, 1009, 1010, 1011, 1019, 1028, 1035, 1037, 1052, 1061, 1062, and 1063

SEQ ID NOS: 157, 161, 163, 168, 197, 200, 205, 218, 221, 242, 250, 331, 334, 340, 412, 413, 414, 419, 425, 452, 453, 454, 456, 459, 462, 491, 591, 618, 622, 663, 665, 667, 670, 672, 690, 694, 695, 709, 759, 760, 761, 762, 767, 819, 820,822, 826, 832, 846, 847, 860, 872, 873, 877, 881, 887, 908, 911, 913, 917, 919, 921, 936, 939, 940, 942, 944, 947, 951, 952, 955, 958, 959, 960, 964, 975, 977, 979, 986, 1023, 1027, 1034, 1038, 1043, 1044, 1049, and 1051;

SEQ ID NOS: 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 332, 333, 334, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 382, 383, 384, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406,

407, 408, 409, 410, 411, 412, 413, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 597, 598, 599, 600, 601, 602, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 629, 630, 631, 632, 633, 634, 635, 636, 637, 639, 640, 641, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 673, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 939, 940, 941, 942, 943, 944, 945, 946, 947, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 1000, 1001, 1002, 1003, 1004, 1005,

1006, 1008, 1009, 1010, 1011, 1012, 1013, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, and 1064; and any and all combinations of the foregoing sets.

The polypeptide translation products of those polynucleotide sequences form sets of preferred olfactory receptor polypeptides, as well as any and all combinations of those polypeptide sets. The preferred sets of polypeptide translation products, and any and all combinations thereof, are also preferred sets for use as libraries of olfactory receptors for scent analysis.

Scent Fingerprinting

It will be appreciated that in many instances, analysis of a scent (whether in terms of receptor primary scent components, receptor quasi-primary scent components, receptor complex scent components, or other scent representations) is of great utility in and of itself, in addition to the utility of that analysis in scent re-creation. Thus, another embodiment of the invention encompasses "scent fingerprinting," which comprises analysis of a scent profile when re-creation of that scent may not be necessary or desirable. The distinction between scent profiling, as defined above, and scent fingerprinting, as defined here, is that scent profiling is a representation of a scent relative to a mammalian olfactory system in such a manner as to provide useful information about the interaction of the scent with that olfactory system, such as sufficient information to enable re-creation of the scent from receptor primary scent components. In contrast, scent fingerprinting can, but does not necessarily, provide such information.

Various applications and examples of scent fingerprinting can include, but are not limited to, the following illustrative situations. Natural gas is widely used as a heating and fuel supply, but is in itself odorless. Utility companies routinely add small amounts of odorants such as mercaptans to allow detection of natural gas leaks in households. Should a leak occur at an unattended site, however, potentially dangerous quantities of natural gas can accumulate. In such areas, a device which can recognize odorants would be useful.

Another use of scent fingerprinting is quality control of a manufacturing process. Many food items, such as freshly-baked bread and pastries, sauces, and cheeses, have distinct

odors. A manufacturer can record a scent fingerprint for a given food item, e.g. spaghetti sauce for packaging in jars. The quality of the product can then be monitored at various stages in manufacture and storage, and deviations from the established scent fingerprint can be used to alert the manufacturer to problems in manufacture or storage. Quality control scent fingerprints are not limited to food items, but can be used in any circumstance where a volatile component of an item of manufacture can be used as a quality control indicator, e.g., perfume, deodorants, solvent mixtures, etc.

While scent fingerprints need not be meaningful in terms of a mammalian olfactory system, it will be readily appreciated that a scent profile, which does represent a scent in a manner relevant to an olfactory system, is a special type of scent fingerprint. Additionally, the response of a device which yields a scent fingerprint of an odor (such as the "artificial nose" described in U.S. Pat. Nos. 5,571,401, 5,698,089, 5,788,833, 5,891,398 and 5,911,872) can be calibrated against the response of a mammalian olfactory system in order to transform the scent fingerprint generated by the device into a true scent profile which can be utilized to re-create an odor using receptor primary scent components, receptor quasi-primary scent components, or receptor complex scent components. The invention encompasses such data transformations.

Scent Editing

Representation of a scent as a scent profile provides the capability of editing the scent. A scent profile which represents a scent in terms of perceptive primary scent components is the most straightforward representation to edit. An example is the perceptive complex primary scent of "burned pizza" comprised of perceptive primary scent components of sausage, cheese, tomato sauce, and burned dough. In order to edit the scent to provide a more pleasant re-creation, the perceptive primary scent component of burned dough would simply be eliminated.

Other scent profiles can be edited using a knowledge of the perception of a particular components. Using our six-receptor example, suppose that the (1, 0, 0, 0, 1, 0) receptor complex scent component is known to provide an unpleasant aspect of the scent, while the (0, 1, 1, 1, 0, 1) component is known to provide the pleasant aspect of the scent. The first complex scent component can be omitted from the edited scent profile, leaving (0, 1, 1, 1, 0, 1) as the edited scent profile. (This would also alter the index values for scent re-creation, from 1 and 1, to 0 and 1.) More complex editing situations can be manipulated using computer algorithms as discussed above.

Individual scent components can be omitted, added, weakened, or intensified, and different scent components can be adjusted in different manners or degrees, depending on the desired result. The editing can be done interactively, with each edited scent emitted by the emitter module for evaluation by the user, or can be done automatically, with removal/weakening or addition/intensifying of particular components specified in advance, on either an absolute scale or relative to other components.

The following examples are presented to illustrate, but not to limit, the invention.

EXAMPLES

Example 1: Isolation of human olfactory receptor cDNAs

Total RNA was extracted from human olfactory epithelium and polyA$^+$ RNA was obtained by oligo-dT selection. This RNA served as template for cDNA synthesis using reagents from the SMART cDNA Library construction kit (Clontech K1051-1; Palo Alto, CA). The Superscript II™ reverse transcriptase (Life Technologies, Gaithersburg, MD) was used for first-strand synthesis.

Double-stranded cDNA was passed through a Chroma-Spin$^+$ STE-100 column (Clontech) to remove unreacted primers and cDNA fragments shorter that 100 nucleotides. The olfactory epithelial cDNA population was then subjected to amplification using primers homologous to conserved regions in GPCRs. The first primer set was homologous to transmembrane segment 2 (TM2) and the second set was homologous to TM 7.5. The TM2 primer set contained 32 oligonucleotides, representing all possible nucleotide sequences capable of encoding the TM2 amino acid sequence motif P-M-Y-F/L-F/Y-F/L, and designed to be non-degenerate at their 3' ends. Sequences of the TM2 primers are as follows:

CCN ATG TAY TTN CTC CTA	SEQ ID NO: 74
CCN ATG TAY TTN CTC CTC	SEQ ID NO: 75
CCN ATG TAY TTN CTC CTG	SEQ ID NO: 76
CCN ATG TAY TTN CTC CTT	SEQ ID NO: 77
CCN ATG TAY TTN CTC TTA	SEQ ID NO: 78
CCN ATG TAY TTN CTC TTC	SEQ ID NO: 79
CCN ATG TAY TTN CTC TTG	SEQ ID NO: 80
CCN ATG TAY TTN CTC TTT	SEQ ID NO: 81
CCN ATG TAY TTN CTT CTA	SEQ ID NO: 82
CCN ATG TAY TTN CTT CTC	SEQ ID NO: 83
CCN ATG TAY TTN CTT CTG	SEQ ID NO: 84

CCN ATG TAY TTN CTT CTT	SEQ ID NO: 85
CCN ATG TAY TTN CTT TTA	SEQ ID NO: 86
CCN ATG TAY TTN CTT TTC	SEQ ID NO: 87
CCN ATG TAY TTN CTT TTG	SEQ ID NO: 88
CCN ATG TAY TTN CTT TTT	SEQ ID NO: 89
CCN ATG TAY TTN TTC CTA	SEQ ID NO: 90
CCN ATG TAY TTN TTC CTC	SEQ ID NO: 91
CCN ATG TAY TTN TTC CTG	SEQ ID NO: 92
CCN ATG TAY TTN TTC CTT	SEQ ID NO: 93
CCN ATG TAY TTN TTC TTA	SEQ ID NO: 94
CCN ATG TAY TTN TTC TTC	SEQ ID NO: 95
CCN ATG TAY TTN TTC TTG	SEQ ID NO: 96
CCN ATG TAY TTN TTC TTT	SEQ ID NO: 97
CCN ATG TAY TTN TTT CTA	SEQ ID NO: 98
CCN ATG TAY TTN TTT CTC	SEQ ID NO: 99
CCN ATG TAY TTN TTT CTG	SEQ ID NO: 100
CCN ATG TAY TTN TTT CTT	SEQ ID NO: 101
CCN ATG TAY TTN TTT TTA	SEQ ID NO: 102
CCN ATG TAY TTN TTT TTC	SEQ ID NO: 103
CCN ATG TAY TTN TTT TTG	SEQ ID NO: 104
CCN ATG TAY TTN TTT TTT	SEQ ID NO: 105

The TM7.5 primer set was designed to contain the reverse complement of all sequences capable of encoding the TM7.5 amino acid sequence motif P-F/L/I/V-I/V-F/Y-S/T-L. The sequences of the TM7.5 primers are as follows:

YYTNGTNYTNRYNCYGATANATNATNGGRTT	SEQ ID NO: 106
YTRTTNCKNAGNWRTANATRAANGGRTT	SEQ ID NO: 107
TCYTTRTTNCKNAGNGWRTANAYNASNGGRTT	SEQ ID NO: 108
TCNTSRTTNCKNARNSARTANATNATNGGRTT	SEQ ID NO: 109
RTTNCKNARNSWRTANATRAANGGRTT	SEQ ID NO: 110

Reagents and enzymes for amplification were from the Advantage cDNA amplification kit (Clontech). A primary amplification reaction was constructed as follows:

5 µl olfactory epithelial cDNA (10-20 µg/ml)

5 µl 10X PCR reaction buffer (Clontech)

1 µl TM2 primer set (10 µM)

1 µl TM7.5 primer set (10 µM)

1 µl dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)

36 µl PCR-grade H_2O

1 µl Advantage polymerase mix (Clontech)

Amplification was conducted in a PE 480 thermal cycler, using 28 cycles of 95°C for 15 sec, 45°C for 45 sec and 72°C for 2 min. After cycling, the amplification mixture was treated for 1 hour at 37°C with 10 Units of BspEI and 10 Units of PstI restriction enzymes, to degrade non-specific amplification products.

The primary amplification products were size-fractionated by agarose gel electrophoresis, and amplification products having a length between 600 and 800 base pairs were selected for secondary amplification.

The secondary amplification reaction was conducted identically to the primary amplification reaction, except that the size-selected primary amplification product was used as template. Secondary amplification reactions containing products which generated a specific gel band of between 600 and 800 base pairs were extracted once with phenol/chloroform and once with chloroform, and nucleic acids were precipitated from the reactions by addition of 0.1 volume of 3M NaOAc (pH 4.8), 20 µg glycogen, and 1.5 volumes of cold 95% ethanol. The precipitate was collected by centrifugation, dried and resuspended in 15 µl distilled water. After the precipitate dissolved, 3 µl loading dye was added, and the sample was subjected to electrophoresis on a 1.0% low-melting agarose gel containing ethidium bromide. Electrophoresis was conducted at 60V for approximately 40 min, with a 1 kb marker in adjoining lanes.

Following electrophoresis, the gel was illuminated with long-wavelength ultraviolet light, and the band was excised from the gel. The gel slice was placed in a 0.5 ml tube, and the tube was heated at 68°C for 15 min. The temperature of the tube was then equilibrated at 45°C. (This is conveniently accomplished in a thermal cycler.) AgarACE™ (Promega) was then added to the tubes, according to the manufacturer's instructions, and incubation at 45°C was continued for 15 min. As a general rule, 2 µl of enzyme per 50 µl of gel slice is adequate. Following AgarACE™ digestion, the digestion mixture was extracted with phenol/chloroform according to the manufacturer's instructions, and nucleic acids were precipitated by addition of 0.1 volume of 3M NaOAc (pH 4.8), 20 µg glycogen, and 1.5 volumes of cold 95% ethanol. The precipitate was collected by centrifugation, dried and resuspended in 5 µl distilled water.

Gel-purified amplification products were cloned using the TOPO XL PCR Cloning Kit (Invitrogen) according to the manufacturer's instructions. After cloning, individual

colonies were selected at random for nucleotide sequence analysis of the inserts, using procedures for sequence determination that are well-known to those of skill in the art.

Example 2: Use of olfactory receptor polypeptides for screening

Components of a scent are identified by determining the interaction between one or more potential odorant molecules and one or more OR polypeptides. For example, if a known original scent involves binding to a particular set of ORs, any subsequent set of molecules which bind to that same set of ORs and stimulate or inhibit the response of the ORs to the same extent as the original scent is capable of re-creating that original scent. If each of the subsequent set of molecules interacts with one and only one OR, then the set of molecules is composed of receptor primary scent components. In similar fashion, scents which involve binding of multiple ORs can be recreated by identifying a molecule, or combination of molecules, which binds to that particular set of ORs.

Binding of molecules to ORs is determined by a number of methods that are well-known in the art including, but not limited to, in vitro and in silico methods as described herein. Binding of molecules to ORs can also be determined or approximated by using quantitative structure-activity relationships as described herein.

Example 3: Identification of agonists and antagonists of olfactory receptors

Interaction of an odorant with a particular OR embedded in the membrane of an olfactory neuron will activate a signaling cascade within the neuron, ultimately resulting in the perception of a particular smell. A molecule, produced for example by combinatorial chemistry, which activates a similar or identical signaling cascade, will induce the perception of the same smell. Such a molecule would be considered a OR agonist. An OR agonist, once identified, can be used as a probe to identify additional agonists, as well as antagonists, of that particular OR.

Assays for the activation and the end product(s) of signaling cascades are known in the art. For example, direct Ca^{++} imaging can be employed, using either dye -labeled Ca^{++} or dyes that are sensitive to Ca^{++} concentration. Such dyes, and techniques for their use, are available from, for example, Molecular Dynamics (Sunnyvale, CA) and Molecular Probes (Eugene, OR).

Because ORs are transmembrane proteins, identification of agonists and/or antagonists for a particular OR require that the OR is present either in a living cell or in a membrane preparation.

In one embodiment of a method for the determination of OR agonists or antagonists, a known OR agonist is labeled *in situ*, or is resynthesized with an attached label, and is bound to an OR. The effect of various test molecules on the binding of the labeled OR agonist is then determined. Labeling of an OR agonist is accomplished by any of a number of methods that are known to those of skill in the art including, but not limited to, various fluorescent labels (for example, chemical fluorochromes or green fluorescent protein). Binding of the OR agonist is measured by any of a number of competitive binding assays, as are known in the art. A test molecule that displaces the agonist from the OR (*i.e.*, reduces the binding of the agonist) is identified as a candidate agonist or antagonist of the particular OR. In a subsequent experiment, the candidate molecule is bound to the OR, and the effect on the signaling cascade induced by the original agonist is determined. A similar of higher level of activation is indicative of an agonist; while a reduced level of activation of the signaling cascade reflects the action of an antagonist.

In additional embodiments of the displacement assay, an unlabeled agonist is used, and its degree of binding is determined by mass spectrometry. *See*, for example, U.S. Patent No. 5,894,063; U.S. Patent No. 5,719,060; and Wei *et al.* (1999) *Nature* **399**:243-246.

In another embodiment, fluorescent microparticles ("beads"), which can be separated by flow cytometry, are used to identify OR agonists and antagonists. Such beads are available, for example, from Luminex (Austin, TX). Multiple different ORs are attached to the beads, wherein each distinct color of bead is associated with a particular OR. The collection of beads, containing different ORs, is exposed to a test molecule or a collection of test molecules, such as can be synthesized by combinatorial chemistry, and binding of the test molecule(s) is determined, for example, by use of a labeled ligand of the test molecule(s). The beads are sorted according to their color by flow cytometry. Correlation of test molecule binding with bead color allows the determination of test molecules capable of binding to the OR. Agonist or antagonist function of an OR binding molecule is determined by methods described *supra*.

Example 4: Summary of search parameters for homology searches

Step 1: (masking) rempolyatmask raw sequence on -NONE- [?] with remAT_moderate (15) . Continue to step 2.

Step 2: (masking) mask masked sequence from step 1 on RepBase [N] with mask_moderate (85) . Continue to step 3.

Step 3: (masking) mask masked sequence from step 2 on VecBase [N] with mask_moderate (85) . Continue to step 4.

Step 4: blastn masked sequence from step 3 on NR-Nuc [N] with blastn_10_hits (V=10 B=10) . If the P/Z score is > 1.0E-50, or no hits are found go to step 5. Otherwise, stop.

Step 5: blastx masked sequence from step 3 on NR-Pro [P] with blastx_10_hits (V=10 B=10) . If the P/Z score is > 1.0E-50, or no hits are found go to step 6. Otherwise, stop.

Step 6: blastn masked sequence from step 3 on GB_CurAwareness-Nuc [N] with blastn_10_hits (V=10 B=10) . If the P/Z score is > 1.0E-50, or no hits are found go to step 7. Otherwise, stop.

Step 7: blastx masked sequence from step 3 on GB_CurAwareness-Pro [P] with blastx_10_hits (V=10 B=10) . If the P/Z score is > 1.0E-50, or no hits are found go to step 8. Otherwise, stop.

Step 8: tblastx masked sequence from step 3 on NR-Nuc [N] with tblastx_10_hits (V=10 B=10) . If the P/Z score is > 1.0E-50, or no hits are found go to step 9. Otherwise, stop.

Step 9: blastn masked sequence from step 3 on EST [N] with blastn_10_hits (V=10 B=10) . If the P/Z score is > 1.0E-50, or no hits are found go to step 10. Otherwise, stop.

Step 10: blastn masked sequence from step 3 on STS [N] with blastn_10_hits (V=10 B=10) . Stop.

Example 5: Summary of search results

Step	Program	Database	Score	Sequences By Best Hit's Score					No Hits	Run	Not Finished	Not Run
1	rempolyat mask	-NONE-[P]	P/Z/E	0	> 1.0 >=	0	>= 1.0 >	0	74	74	0	0
2	mask	RepBase[N]	P/Z/E	0	> 1.0 >=	0	>= 1.0 >	0	74	74	0	0
3	mask	VecBase[N]	P/Z/E	0	> 1.0 >=	0	>= 1.0 >	0	74	74	0	0
4	blastn	NR-Nuc[N]	P/Z/E	46	< 1.0E-20 <=	28			0	74	0	0
5	blastx	NR-Pro[P]	P/Z/E	16	< 1.0E-20 <=	34			0	50	0	24
6	blastn	GB_CurAwareness-Nuc[N]	P/Z/E	17	< 1.0E-20 <=	31			0	48	0	26
7	blastx	GB_CurAwareness-Pro[P]	P/Z/E	13	< 1.0E-20 <=	28			2	43	0	31
8	tblastx	NR-Nuc[N]	P/Z/E	14	< 1.0E-20 <=	29			0	43	0	31
9	blastn	EST[N]	P/Z/E	10	< 1.0E-20 <=	33			0	43	0	31
10	blastn	STS[N]	P/Z/E	5	< 1.0E-20 <=	33			0	38		

Example 6. Datamining and analysis from GenBank

Datamining. A datamining pipeline was built to detect all available OR-like sequences in the public databases and to update the results as new database versions are released. tblastn (Altschul et al., 1997) was used to compare amino acid query sequences to the non-redundant version of GenBank (partitions nt, htg and est_human, all updated to August 6th, 2000), with a non-stringent expectation value cutoff of 1e-4. The queries used included 96 curated OR sequences representing all known families (SEQ ID NO:2651 through SEQ ID NO:2747) and 249 additional HORDE entries (SEQ ID NO:2402 through SEQ ID NO:2650). In a second round 105 newly mined mouse genes (SEQ ID NO:2296 through SEQ ID NO:2401) and 344 newly mined human genes (SEQ ID NO:2009 through SEQ ID NO:2295) were used as additional queries (all datasets are available

electronically). All resulting database entries were catalogued by species and subdivided into four types: mRNA, EST, DNA and genomic, the latter including entries annotated with keyword HTGS_PHASE1-3, or with length at least 10 kb. Low-pass genomic sampling sequences were ignored (keyword HTGS_PHASE0). In addition, a set of 132 olfactory sequence tag (OST) sequences was used . All sequences used were split into contigs according to annotation or, where unavailable, according to runs of at least 50 Ns. All resulting contigs were analyzed for interspersed repeats using RepeatMasker (Smit and Green, 1997). Subcontigs were defined as segments between interspersed repeats, ignoring simple repeats and low-complexity regions.

Localization of genomic clones. The University of Santa Cruz (UCSC) Working Draft Sequence ("golden path", http://genome.ucsc.edu) presents a first tentative assembly of the finished and draft human genomic sequence based on the WUSTL clone map (http://genome.wustl.edu/gsc). The "golden path" data was used to assign a coordinate to each finished or unfinished genomic clone, in Mb from the p telomere. In parallel, the Unified DataBase (UDB) was used to assign similar Mb coordinates to the clones, based on their marker contents (Chalifa-Caspi et al., 1998). The two maps are largely colinear, and were integrated based on the coordinates of clones that could be localized in both. Clones for which no coordinate could be obtained by either method were assigned a chromosome according to UDB, by sequence similarity to another mapped clone, by annotation, or by e-PCR (Schuler, 1997).

Detection of OR sequences. Each subcontig was compared using FASTY (Pearson et al., 1997) to a curated set of OR protein sequences from several species, yielding a conceptual translation product. The possibility of a pseudogene being disrupted by the insertion of interspersed repeats was taken into account, with the two or more resulting parts being therefore located in different subcontigs. Such compatible candidate sequences were automatically joined into a combined reconstructed pseudogene. Whenever possible, all resulting sequences were trimmed or extended to use a suitable ATG codon for initiation and to end at a stop codon, but avoiding those stop codons that yield products shorter than 275 amino acids. The sequences were finally split into OR or non-OR by comparing them to previously recognized OR sequences and to a non-redundant database of non-OR GPCRs which we extracted from Swiss-Prot. To be automatically classified as an OR, a

new sequence has to be at least 40% identical over at least 100 amino acids to another OR. A more stringent cutoff (50%) was required for shorter sequences.

Definition of OR genes. A given gene could be represented in more than one overlapping genomic clone. Such redundancy was removed by considering two sequences as representing the same gene, if they are in the same chromosome, located in clones less than 300 kb apart and at least 99% identical at the nucleotide level. An exception to this rule is when two genes coappear in the same clone, in which case they were considered to be distinct genes. Sequences localized to a chromosome but without a coordinate were only compared to other sequences within that chromosome, and finally those sequences lacking a chromosomal assignment were compared to the rest, applying only the criterion of sequence similarity. For each resulting gene with more than one constituent sequence, a consensus nucleotide sequence was created after multiple alignment by ClustalW (Higgins et al., 1996) using the fast comparison parameter. This was followed by conceptual translation and end trimming to suitable start and stop codons, as above. Genes with length at least 275 amino acids without frame disruptions (frameshifts, in-frame stop codons or disrupting interspersed repeats) were considered to be full-length and apparently intact. For partial sequences without frame disruptions no statement could be made on their apparent functionality, except when the partial sequences were observed in the genome as such, in which case they were considered to be pseudogenes. Finally, each OR gene was assigned a family and subfamily by amino acid sequence similarity to previously classified OR genes.

The references cited in this example are: Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-402; Chalifa-Caspi, V., Prilusky, J. and Lancet, D. (1998) The Unified Database. Weizmann Institute of Science, Bioinformatics Unit and Genome Center (Rehovot, Israel). World Wide Web URL: bioinformatics.weizmann.ac.il/udb; Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266: 383-402; Pearson, W. R., Wood, T., Zhang, Z. and Miller, W. (1997) Comparison of DNA sequences with protein sequences. Genomics 46: 24-36; Schuler, G. D. (1997) Sequence mapping by electronic PCR. Genome Res 7: 541 50; and Smit, A. F.

A. and Green, P. (1997) RepeatMasker at URL: repeatmasker.genome.washington.edu/cgi-bin/RM2_req.pl.

Tables 1 and 2 contain additional information regarding SEQ ID NO. 153 to SEQ ID NO. 1085. The explanation of the entries in Tables 1 and 2 is as follows:

Symbol: The Human Genome Organization gene symbol, as allotted by a procedure to be published soon. OR = Olfactory Receptor, numeral to the immediate right - family designation, capital letters - subfamily designation, rightmost numeral - individual gene within subfamily, n appearing when such number is not assigned yet; P = Pseudogene.

All ORs within a family share at least 40% protein sequence identity.

All ORs within a subfamily share at least 60% protein sequence identity.

HORDE: The H serial number within the Human Olfactory Receptor Data Exploratorium (URL bioinfo.weizmann.ac.il/HORDE). The numeral 38 represents the HORDE build (version), gxxx is the individual gene number.

Digi: Appearance of a DSnn serial number here means that the sequence has been PCR-amplified from human olfactory epithelial cDNA using degenerate primers at the transmembrane helix 2 and transmembrane helix 7. See separate page for explanations on the analysis of the DS entries.

OST: OSTnnn is the serial number of the sequence in the Olfactory Sequence Tag collection in the Lancet laboratory (URL bioinfo.weizmann.ac.il/HORDE). Appearance here means that the sequence has been PCR-amplified from human genomic DNA using degenerate primers at the transmembrane helix 2 and transmembrane helix 7. There are a total of 112 OST sequences.

Trivial name: One or more aliases given to the same gene by different laboratories. Many of the trivial names are of the form ORnn-xx, whereby nn is a chromosome number and xx is an arbitrary numerical identifier.

Tran: (transcribed) Plus appears if the entry was sequenced from cDNA, or was found in the Expressed Sequence Tags (EST) databases. Plus also appears if in the public databases the gene was annotated as mRNA.

Int.: (intact) "Yes" indicates that the gene may be intact, as there are no obvious sequence frame disruptions. "Put" (putative) indicates the same, except that the known sequence is short, hence there may be disruptions in the unsequenced segments. “Pol”

indicates a polymorphism between intact and pseudogenic alleles. When no word appears, this indicates a pseudogene.

E: (Extent) FL indicates that the Full Length sequence is known (typically 310 ± 30 amino acids).

D: The number of sequence disruptions in the known sequence of a pseudogene.

C: The human chromosomal location of the OR gene, assigned as described under Mb coord.

Mb coord: The location of the OR gene within a human chromosome, in magabase units, beginning at the p-telomere and ending at the q-telomere, computed based on integrating information from Unified Database (URL is bioinfo.weizmann.ac.il/udb) and the University of California Santa Cruz (URL is genome.ucsc.edu).

CDR: The 17 amino acids suggested to line the odorant ligand binding pocket, delineated by the extracellular 2/3 of transmembrane helices 3,4 and 5. The assignment is based on an algorithm at URL bioinformatics.weizman.ac.il/HORDE/humanGenes/CDR.html.

%: (% id) The percent protein identity between the human sequence in the current line and the known rodent (rat or mouse) OR sequence to which it bears the highest similarity.

S: (Species) Rat (R)or mouse (M).

Acc: The Genbank accession number of the clone that contains the rodent sequence.

Range: The positions x ... y of the first and last bases within the rodent which constitute the OR coding region. If x>y then the OR is on the reverse strand.

Table 1

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
153	OR10D3	H38g001			HSHTPCRX09			
154	OR7EnP	H38g002						FL
155	OR1D5	H38g003		OST901	OR17-31	+	pol	FL
156	OR10NnP	H38g00						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
		4						
157	OR2F1	H38g00 5		OST902	OLF3;OR7-139;OR7-140	+	yes	FL
158	OR7EnP	H38g00 6						FL
159	OR8FnP	H38g00 7						FL
160	OR2Q1P	H38g00 8			DJ0669B10;OR7-2			FL
161	OR2W1	H38g00 9			AL035402-B;dJ88J8.1;hs6M1-15		yes	FL
162	OR7EnP	H38g01 0				+		FL
163	OR6B1	H38g01 1	DS119		OR7-3;WUGSC:H_DJ0669B10.3	+	yes	FL
164	OR10Kn	H38g01 2					yes	FL
165	ORnP	H38g01 3				+		FL
166	OR4F2P	H38g01 4			HS191N21;dJ191N21.4;hs6M1-11			FL
167	OR7EnP	H38g01 5						FL
168	OR1F2P	H38g01 6			OLFMF2	+	yes	FL
169	OR2P1P	H38g01 7			AL035402-A;dJ88J8.2;hs6M1-26			
170	OR7E43P	H38g01 8		OST903	OR4-116			FL
171	OR4F1	H38g01 9			HSDJ0609N19			FL
172	OR7E55P	H38g02 0		OST904	OR2DG;OR3.2			FL
173	OR13Dn	H38g02 1					yes	FL
174	OR4CnP	H38g02 2						FL
175	OR10D1P	H38g02 3		OST074	HSHTPCRX03	+		FL
176	OR4Cn	H38g02					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
		4						
177	OR8GnP	H38g02 5						
178	OR13CnP	H38g02 6						FL
179	OR4CnP	H38g02 7						FL
180	OR13Cn	H38g02 8					yes	FL
181	OR4CnP	H38g02 9						
182	OR51Bn	H38g03 0					yes	FL
183	OR7E5P	H38g03 1		OST905	OR11-12			FL
184	OR13Cn	H38g03 2					yes	FL
185	OR4Sn	H38g03 3					yes	FL
186	OR51BnP	H38g03 4						FL
187	OR6JnP	H38g03 5						FL
188	OR51Bn	H38g03 6					yes	FL
189	OR7EnP	H38g03 7						FL
190	OR2An	H38g03 8					yes	FL
191	OR7E22P	H38g03 9			OR3.6;OR6DG			FL
192	OR7E4P	H38g04 0			OR11-11a			FL
193	OR7E66P	H38g04 1		OST906	OR3.3;OR3DG;hg630			FL
194	OR6Mn	H38g04 2					yes	FL
195	OR2ALnP	H38g04 3						
196	OR6MnP	H38g04 4						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
197	OR4D1	H38g045			AC005962-A;HSTPCR16	+	yes	FL
198	OR5D2P	H38g046		OST907	OR11-7a;OR912-91			FL
199	OR7E38P	H38g047		OST127	AC004967	+		FL
200	OR4D2	H38g048			AC005962-B		yes	FL
201	OR7E7P	H38g049			AC004967-A			FL
202	OR5AHnP	H38g050						
203	OR2U2P	H38g051			AL050339-B;dJ974I11.2;hs6M1-23			FL
204	OR2U1P	H38g052			974I11;AL050339-C;dJ974I11.3;hs6M1-24			FL
205	OR2H2	H38g053			AC006137-A;dJ271M21.2;hs6M1-12		yes	FL
206	OR2H5P	H38g054		OST616	HS271M21;hs6M1-13			FL
207	OR2In	H38g055				+	yes	FL
208	OR11HnP	H38g056						FL
209	OR7EnP	H38g057				+		
210	OR9In	H38g058					yes	FL
211	OR2AFnP	H38g059						FL
212	OR13KnP	H38g061						FL
213	OR13Cn	H38g062					yes	FL
214	OR13Fn	H38g063					yes	FL
215	OR9Qn	H38g064					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
216	OR2TnP	H38g065						FL
217	OR4Kn	H38g066					yes	FL
218	OR2B8P	H38g067			dJ313I6.4;hs6M1-29P		yes	FL
219	OR2Tn	H38g068					yes	FL
220	OR4Kn	H38g069					yes	FL
221	OR2A4	H38g070			WUGSC:H_DJ0988G15.2	+	yes	FL
222	OR7EnP	H38g071						FL
223	OR4Kn	H38g072					yes	FL
224	OR13InP	H38g073						FL
225	OR7EnP	H38g074						FL
226	OR6Jn	H38g075					yes	FL
227	OR4Mn	H38g076					yes	FL
228	OR4VnP	H38g077						FL
229	OR6Xn	H38g078					yes	FL
230	OR51Gn	H38g079					yes	FL
231	OR6EnP	H38g080						FL
232	OR4NnP	H38g081						FL
233	OR6MnP	H38g082						FL
234	OR4Nn	H38g083					yes	FL
235	OR4Cn	H38g084					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
236	OR4KnP	H38g085						FL
237	ORnP	H38g086						
238	OR5D3	H38g087		OST908	OR11-8b;OR11-8c			
239	OR2G1P	H38g088	DS13;DS16	OST619	dJ974I11.4;hs6M1-25	+		FL
240	OR4Kn	H38g089					yes	FL
241	OR8BnP	H38g090						FL
242	OR2B2	H38g091			OR6-1;dJ193B12.4		yes	FL
243	OR7EnP	H38g092						FL
244	OR4KnP	H38g093						FL
245	OR2AD1P	H38g094			dJ25J6.1;hs6M1-8P			FL
246	OR1AAnP	H38g095						FL
247	OR1E3P	H38g096			OR17-210			FL
248	OR8BnP	H38g097						FL
249	OR5Hn	H38g098					yes	FL
250	OR1G1	H38g099		OST909	OR17-130;OR17-209	+	yes	FL
251	OR5HnP	H38g100						FL
252	ORnP	H38g101						
253	ORnP	H38g102						
254	OR4PnP	H38g103						FL
255	OR13Hn	H38g104					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
256	OR7D1P	H38g105		OST910	CIT-B-440L2;OR19-131;OR19-A			FL
257	OR4KnP	H38g106						FL
258	OR7E24	H38g107		OST911	CIT-B-440L2;OR19-8	+		FL
259	OR51NnP	H38g108						FL
260	OR7E18P	H38g109		OST912	OR19-14;TPCR26	+		FL
261	OR7E19P	H38g110		OST913	HSCIT-B-440L2;OR19-7;TPCR110	+		FL
262	OR7E41P	H38g111		OST914	OR11-20;hg84			FL
263	OR2R1	H38g112		OST058				FL
264	OR10ACnP	H38g113						FL
265	OR51Ln	H38g114					yes	FL
266	OR52JnP	H38g115						FL
267	OR9LnP	H38g116						
268	OR51PnP	H38g117						FL
269	OR5HnP	H38g118						FL
270	OR51An	H38g119					yes	FL
271	OR5HnP	H38g120						FL
272	ORnP	H38g121						
273	OR52En	H38g122					yes	FL
274	OR5Hn	H38g123					yes	FL
275	OR4CnP	H38g124						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
276	OR52En	H38g125					yes	FL
277	OR10Dn	H38g126					yes	FL
278	OR5HnP	H38g127						FL
279	OR13An	H38g128					yes	FL
280	OR5HnP	H38g129						FL
281	OR5Kn	H38g130					yes	FL
282	OR7EnP	H38g131						FL
283	OR4DnP	H38g132						FL
284	OR2ARnP	H38g133						
285	OR7E29P	H38g134		OST032				FL
286	OR4CnP	H38g135						FL
287	OR5PnP	H38g136						FL
288	OR7EnP	H38g137						FL
289	OR56An	H38g138					yes	FL
290	OR56AnP	H38g139						
291	OR5Pn	H38g140					yes	FL
292	OR7E53P	H38g141		OST915	OR3-142;OR3-143			FL
293	OR5Pn	H38g142					yes	FL
294	OR52Ln	H38g143					yes	FL
295	OR5E1	H38g144			HSTPCR24	+		FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
296	OR56AnP	H38g145						
297	OR4KnP	H38g146						
298	OR52Ln	H38g147					yes	FL
299	OR7EnP	H38g148						
300	OR52XnP	H38g149						FL
301	ORnP	H38g150						
302	OR56An	H38g151					yes	FL
303	OR56AnP	H38g152						
304	OR1R1P	H38g153			OR17-1			FL
305	OR52EnP	H38g154						FL
306	OR51AnP	H38g155						FL
307	OR51An	H38g156					yes	FL
308	OR4CnP	H38g157						FL
309	OR52JnP	H38g158						FL
310	OR4RnP	H38g159						
311	OR52Jn	H38g160					yes	FL
312	OR4CnP	H38g161						FL
313	OR51AnP	H38g162						FL
314	OR7EnP	H38g163						FL
315	OR5MnP	H38g164						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
316	OR10ABnP	H38g165						FL
317	OR52SnP	H38g166						FL
318	OR5Mn	H38g167					yes	FL
319	OR10Sn	H38g168					yes	FL
320	OR5MnP	H38g169						FL
321	OR10Gn	H38g170					yes	FL
322	ORnP	H38g171						FL
323	OR5MnP	H38g172						FL
324	OR10GnP	H38g173						
325	OR10TnP	H38g174						FL
326	ORnP	H38g175						
327	OR10RnP	H38g176						FL
328	OR5MnP	H38g177						FL
329	OR7EnP	H38g178						FL
330	OR10Tn	H38g179					yes	FL
331	OR1E1	H38g180	DS37;DS43;DS46	OST916	HGMP07I;OR17-2;OR17-32	+	yes	FL
332	OR5BKnP	H38g181						
333	OR5MnP	H38g182						FL
334	OR3A3	H38g183		OST917	OR17-137;OR17-16;OR17-201	+	yes	FL
335	OR10ADnP	H38g184	DS10			+		FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
336	OR10Rn	H38g185				+	yes	FL
337	OR5TnP	H38g186						FL
338	OR4GnP	H38g187						FL
339	OR6Yn	H38g188					yes	FL
340	OR1E2	H38g189		OST918	OR17-135;OR17-93	+	yes	FL
341	OR8Hn	H38g190					yes	FL
342	OR4Fn	H38g191					yes	FL
343	OR10Kn	H38g192					yes	FL
344	OR7LnP	H38g193						
345	OR8InP	H38g194						FL
346	OR10RnP	H38g195						
347	OR2AFnP	H38g196						FL
348	OR8Kn	H38g197					yes	FL
349	ORnP	H38g198						
350	OR8KnP	H38g199						FL
351	OR51Hn	H38g200					yes	FL
352	OR7EnP	H38g201						FL
353	ORnP	H38g202						
354	OR5BMnP	H38g203						FL
355	OR10GnP	H38g204						

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
356	OR2Yn	H38g205					yes	FL
357	OR10DnP	H38g206						FL
358	OR3BnP	H38g207						FL
359	OR8Dn	H38g208					yes	FL
360	OR5RnP	H38g209						
361	OR10Gn	H38g210					yes	FL
362	OR5BDnP	H38g211						FL
363	OR5ALnP	H38g212						FL
364	OR52HnP	H38g213						
365	OR10Gn	H38g214					yes	FL
366	OR5Mn	H38g215					yes	FL
367	OR51Mn	H38g216					yes	FL
368	OR6Tn	H38g217	DS15;DS146;DS147			+	yes	FL
369	OR6DnP	H38g218						FL
370	OR4B1	H38g219		OST208			yes	FL
371	OR5ALnP	H38g220						FL
372	OR51Qn	H38g221					yes	FL
373	OR4Dn	H38g222					yes	FL
374	OR52Nn	H38g223					yes	FL
375	OR4Xn	H38g224					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
376	OR8Jn	H38g225					yes	FL
377	OR51JnP	H38g226						FL
378	OR10Gn	H38g227					yes	FL
379	OR52En	H38g228					yes	FL
380	OR4Xn	H38g229					yes	FL
381	OR10A2	H38g230	DS5;DS53;DS56	OST363		+		FL
382	OR5Mn	H38g231					yes	FL
383	OR52En	H38g232					yes	FL
384	OR8Kn	H38g233					yes	FL
385	OR10An	H38g234	DS55			+	yes	FL
386	OR8LnP	H38g235						FL
387	OR5BPnP	H38g236						
388	OR52Nn	H38g237					yes	FL
389	ORnP	H38g238						
390	OR8JnP	H38g239						FL
391	OR5Mn	H38g240					yes	FL
392	OR52En	H38g241					yes	FL
393	OR5Tn	H38g242					yes	FL
394	OR52NnP	H38g243						FL
395	OR4B2P	H38g244		OST919	hg449			FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
396	OR51KnP	H38g245						FL
397	OR52QnP	H38g246						FL
398	OR4Fn	H38g247					yes	FL
399	OR11MnP	H38g248						
400	OR52Nn	H38g249					yes	FL
401	OR56An	H38g250					yes	FL
402	OR5AWnP	H38g251						FL
403	OR52Nn	H38g252					yes	FL
404	ORnP	H38g253						
405	OR52EnP	H38g254						FL
406	OR5BHnP	H38g255						FL
407	OR4QnP	H38g256						FL
408	OR51En	H38g257					yes	FL
409	OR11KnP	H38g258						FL
410	OR12D1P	H38g259			AC004174-B;dJ994E9.7;hs6M1-19			FL
411	OR4NnP	H38g260				+		FL
412	OR11A1	H38g261			AC004174-A;dJ994E9.6;hs6M1-18	+	yes	FL
413	OR10C1	H38g262			AC004174;dJ994E9.5;hs6M1-17	+	yes	FL
414	OR2H1	H38g263	DS114		OLFR42A-9004-14;OR6-2;dJ994E9.4;hs6M1-16	+	yes	FL
415	OR9RnP	H38g264						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
416	OR4FnP	H38g26 5						
417	OR7D4	H38g26 6		OST920	OR19-B;hg105			FL
418	OR7E25P	H38g26 7		OST921	CIT-B-440L2;OR19-C			FL
419	OR2D2	H38g26 8			OR11-610		yes	FL
420	OR10An	H38g26 9					yes	FL
421	OR2WnP	H38g27 0				+		
422	OR7E16P	H38g27 1		OST922	CIT-B-440L2;OR19-133;OR19-9			FL
423	OR52Pn	H38g27 2					yes	FL
424	OR6AnP	H38g27 3						FL
425	OR7D2	H38g27 4	DS70;DS73	OST923	HTPCRH03;OR19-4	+	yes	FL
426	OR52UnP	H38g27 5						FL
427	OR2AGn	H38g27 6					yes	FL
428	OR7G3	H38g27 7		OST085			yes	FL
429	OR56BnP	H38g27 8						FL
430	OR2AGnP	H38g27 9						FL
431	OR56Bn	H38g28 0					yes	FL
432	OR6AnP	H38g28 1						FL
433	OR4FnP	H38g28 2						FL
434	OR6Wn	H38g28 3					yes	FL
435	OR4Mn	H38g28 4					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
436	OR52YnP	H38g285						
437	OR11HnP	H38g286						FL
438	OR9An	H38g287					yes	FL
439	OR5Mn	H38g288					yes	FL
440	OR6Vn	H38g289					yes	FL
441	OR4Nn	H38g290				+	yes	FL
442	OR51AnP	H38g291						FL
443	OR9PnP	H38g292						
444	OR4H6P	H38g293			OR15-71;OR15-82			FL
445	OR51FnP	H38g294						FL
446	OR7E1P	H38g295			AC004923			FL
447	OR51Tn	H38g296					yes	FL
448	OR2Vn	H38g297					yes	FL
449	OR51HnP	H38g298						FL
450	OR51An	H38g299					yes	FL
451	OR2AInP	H38g300						FL
452	OR2F2	H38g301			OR7-1;WUGSC:H_DJ0669B10.1		yes	FL
453	OR1F12	H38g302			dJ313I6.5;hs6M1-35P		yes	FL
454	OR7G1P	H38g303			OR19-15		yes	FL
455	OR7G2	H38g304		OST260			yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
456	OR1M1	H38g305		OST924	OR19-6		yes	FL
457	OR51UnP	H38g306						
458	OR52Hn	H38g307					yes	FL
459	OR1F1	H38g308		OST925	OLFMF;OR16-36;OR16-37;OR16-88;OR16-89;OR16-90	+	yes	FL
460	OR10PnP	H38g309						
461	OR4FnP	H38g310						FL
462	OR2T1	H38g311			OR1-25		yes	FL
463	OR7EnP	H38g312						FL
464	OR51Gn	H38g313					yes	FL
465	OR2Tn	H38g314					yes	FL
466	OR5BGnP	H38g315						
467	OR5WnP	H38g316						FL
468	OR51Sn	H38g317					yes	FL
469	OR5WnP	H38g318						
470	OR51AnP	H38g319						FL
471	OR5Dn	H38g320					yes	FL
472	OR7EnP	H38g321						FL
473	OR51Fn	H38g322					yes	FL
474	OR5Dn	H38g323					yes	FL
475	OR52Rn	H38g324					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
476	ORnP	H38g325						FL
477	OR7EnP	H38g326						FL
478	OR6Qn	H38g327					yes	FL
479	OR4Fn	H38g328					yes	FL
480	OR7EnP	H38g329						
481	OR7En	H38g330					yes	FL
482	OR4Nn	H38g331					yes	FL
483	OR2ASnP	H38g332						
484	OR11Hn	H38g333					yes	FL
485	OR2Tn	H38g334					yes	FL
486	OR2TnP	H38g335						
487	OR2AKnP	H38g336						FL
488	ORnP	H38g337						
489	OR5DnP	H38g338						FL
490	OR7EnP	H38g339						
491	OR5L2	H38g340			HSHTPCRX16	+	yes	FL
492	OR5Dn	H38g341					yes	FL
493	ORnP	H38g342						
494	OR10Qn	H38g343					yes	FL
495	OR9MnP	H38g344						

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
496	OR7E62P	H38g345		OST926	OR2-4;OR2-52;OR2-53;OR2-75			FL
497	OR9LnP	H38g346						FL
498	OR7E46P	H38g347		OST379				FL
499	OR1S1	H38g348		OST034			yes	FL
500	OR5DnP	H38g349						
501	OR9InP	H38g350						FL
502	OR5Dn	H38g351					yes	FL
503	OR9QnP	H38g352						FL
504	OR51CnP	H38g353						
505	OR5WnP	H38g354						
506	OR9InP	H38g355						FL
507	OR51AnP	H38g356						FL
508	OR5L1	H38g357		OST262			yes	FL
509	OR7EnP	H38g358				+		
510	OR5BLnP	H38g359						
511	OR51En	H38g360					yes	FL
512	OR51Dn	H38g361					yes	FL
513	OR52In	H38g362					yes	FL
514	OR4KnP	H38g363	DS67			+		FL
515	OR52In	H38g364					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
516	OR4KnP	H38g365						FL
517	OR52MnP	H38g366						FL
518	ORnP	H38g367						
519	ORnP	H38g368						
520	ORnP	H38g369						FL
521	ORnP	H38g370						
522	ORnP	H38g371						
523	ORnP	H38g372						
524	ORnP	H38g373						
525	ORnP	H38g374						
526	OR6Pn	H38g375					yes	FL
527	OR7EnP	H38g376						FL
528	ORnP	H38g377						
529	OR7EnP	H38g378						FL
530	ORnP	H38g379						
531	OR10XnP	H38g380						FL
532	OR10Zn	H38g381					yes	FL
533	OR6KnP	H38g382						FL
534	OR6Kn	H38g383					yes	FL
535	OR1FnP	H38g384						

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
536	OR1ABnP	H38g385						
537	OR52MnP	H38g386						FL
538	OR1XnP	H38g387						FL
539	OR4FnP	H38g388						
540	OR52MnP	H38g389						FL
541	OR2Vn	H38g390					yes	FL
542	OR2V1P	H38g391		OST265				FL
543	OR2Zn	H38g392					yes	FL
544	OR52KnP	H38g393				+		
545	OR10Hn	H38g394					yes	FL
546	OR2Dn	H38g395					yes	FL
547	OR7EnP	H38g396						
548	OR11GnP	H38g397						FL
549	ORnP	H38g398						
550	OR11Gn	H38g399					yes	FL
551	OR11HnP	H38g400						FL
552	OR6Kn	H38g401					yes	FL
553	OR11Hn	H38g402					yes	FL
554	OR6KnP	H38g403						
555	OR11HnP	H38g404						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
556	OR6KnP	H38g405						FL
557	OR6Kn	H38g406					yes	FL
558	OR2Ln	H38g407					yes	FL
559	OR4GnP	H38g408						
560	OR6Nn	H38g409					yes	FL
561	OR2LnP	H38g410						
562	OR9A1	H38g411			HSHTPCRX06			
563	OR6Nn	H38g412					yes	FL
564	OR10Hn	H38g413					yes	FL
565	OR7EnP	H38g414						FL
566	OR2AQnP	H38g415						
567	OR2LnP	H38g416						FL
568	OR5ARn	H38g417					yes	FL
569	OR7EnP	H38g418						FL
570	OR10AAnP	H38g419						FL
571	OR10JnP	H38g420						FL
572	OR5A1P	H38g421	DS69;DS71;DS128;DS129	OST181		+	yes	FL
573	OR2AHnP	H38g422						FL
574	OR10JnP	H38g423						FL
575	OR56BnP	H38g42						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
		4						
576	OR5M1	H38g425		OST050			yes	FL
577	OR52WnP	H38g426						
578	OR5AMnP	H38g427						FL
579	OR52BnP	H38g428						FL
580	OR5MnP	H38g429						FL
581	OR5APnP	H38g430						FL
582	OR56Bn	H38g431					yes	FL
583	OR5APn	H38g432					yes	FL
584	OR52Bn	H38g433					yes	FL
585	OR9Gn	H38g434					yes	FL
586	OR52Kn	H38g435					yes	FL
587	OR5MnP	H38g436						FL
588	OR52Kn	H38g437					yes	FL
589	OR52KnP	H38g438				+		FL
590	OR52BnP	H38g439						FL
591	OR2B6P	H38g440			OR6-31		yes	FL
592	OR2WnP	H38g441						FL
593	OR2AnP	H38g442						FL
594	ORnP	H38g443						
595	OR2LnP	H38g444						

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
596	OR2W2P	H38g445	DS148		dJ313I6.2;hs6M1-30P	+		FL
597	OR2LnP	H38g446						
598	OR2B7P	H38g447			dJ313I6.3;hs6M1-31P			FL
599	OR2Ln	H38g448					yes	FL
600	OR5BFn	H38g449					yes	FL
601	OR2LnP	H38g450						FL
602	OR7EnP	H38g451						
603	OR1H1	H38g452	DS122	OST26		+		FL
604	ORnP	H38g453						
605	OR4Dn	H38g454					yes	FL
606	OR1Ln	H38g455					yes	FL
607	OR5AXn	H38g456					yes	FL
608	OR5An	H38g457					yes	FL
609	OR5AYn	H38g458					yes	FL
610	OR13Gn	H38g459					yes	FL
611	OR5BBnP	H38g460						
612	OR9GnP	H38g461						FL
613	OR2TnP	H38g462						FL
614	ORnP	H38g463						FL
615	OR1Jn	H38g464				+	yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
616	OR2CnP	H38g465						FL
617	OR9GnP	H38g466						FL
618	OR2C1	H38g467			OLFmf3	+	yes	FL
619	OR51AnP	H38g468						
620	OR9Gn	H38g469					yes	FL
621	OR52Bn	H38g470					yes	FL
622	OR1K1	H38g471			hg99		yes	FL
623	OR51RnP	H38g472						FL
624	OR7EnP	H38g473						FL
625	OR52PnP	H38g474						FL
626	OR7EnP	H38g475						FL
627	OR7EnP	H38g476						
628	OR4KnP	H38g477	DS66		OR21-1	+		FL
629	OR4KnP	H38g478			OR21-2			FL
630	OR7EnP	H38g479						
631	OR51In	H38g480					yes	FL
632	OR51In	H38g481					yes	FL
633	OR2AnP	H38g482						
634	OR2A2	H38g483		OST008				FL
635	OR2AnP	H38g484						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
636	OR2Gn	H38g485					yes	FL
637	OR2AnP	H38g486						
638	OR6Fn	H38g487	DS20;DS21;DS23;DS27;DS28;DS39;DS40;DS113;DS126;DS135;DS137;DS138;DS139;DS140;DS141;DS145			+	yes	FL
639	OR2AnP	H38g488						
640	OR2Gn	H38g489					yes	FL
641	OR7E37P	H38g490			hg533	+		FL
642	OR5AVn	H38g491	DS4;DS6;DS11			+	yes	FL
643	OR2AJnP	H38g492						FL
644	OR13EnP	H38g493						FL
645	OR2Cn	H38g494					yes	FL
646	OR2TnP	H38g495						
647	OR2WnP	H38g496						
648	OR13Jn	H38g497					yes	FL
649	OR6RnP	H38g498						FL
650	OR5ATn	H38g499					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
651	OR2Zn	H38g500					yes	FL
652	OR4Ln	H38g501					yes	FL
653	OR4UnP	H38g502						FL
654	OR4Fn	H38g503					yes	FL
655	OR4FnP	H38g504						FL
656	OR4Fn	H38g505					yes	FL
657	OR4Fn	H38g506					yes	FL
658	OR4AnP	H38g507						FL
659	OR4LnP	H38g508						FL
660	OR7E33P	H38g509		OST927	hg688			FL
661	OR2Cn	H38g510					yes	FL
662	OR4Kn	H38g511					yes	FL
663	OR5U1	H38g512			bA150A6.4;hs6M1-28		yes	FL
664	OR4Kn	H38g513					yes	FL
665	OR5V1	H38g514			bA150A6.2;hs6M1-21		yes	FL
666	OR4QnP	H38g515						FL
667	OR12D3	H38g516			bA150A6.1;hs6M1-27		yes	FL
668	OR4Kn	H38g517					yes	FL
669	OR51CnP	H38g518						
670	OR1J2	H38g519		OST044	hg152		yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
671	OR5BJnP	H38g520						
672	OR1J1	H38g521	DS130	OST928	hg32	+	yes	FL
673	OR13En	H38g522					put	
674	OR4KnP	H38g523	DS1			+		FL
675	OR1LnP	H38g524						
676	OR2CnP	H38g525						
677	OR4TnP	H38g526						FL
678	OR5BnP	H38g527						
679	OR4Kn	H38g528					yes	FL
680	OR11Ln	H38g529					yes	FL
681	OR7E68P	H38g530		OST929	OR912-108;OR912-109;OR912-110;OR912-46;hg523;hg674			FL
682	OR7EnP	H38g531						FL
683	OR7E31P	H38g532		OST016;OST205				FL
684	OR7EnP	H38g533						FL
685	OR5AKnP	H38g534						FL
686	OR5AKn	H38g535					yes	FL
687	OR5AKn	H38g536					yes	FL
688	OR5BQnP	H38g537						
689	OR1Nn	H38g538	DS136;DS142			+	yes	FL
690	OR1J4	H38g539		OST930	HSHTPCRX01	+	yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
691	OR1Nn	H38g540					yes	FL
692	OR2AnP	H38g541						FL
693	OR2ANnP	H38g542						
694	OR5K1	H38g543			HSHTPCRX10	+	yes	FL
695	OR2K2	H38g544			HSHTPCRH06		yes	FL
696	OR8Hn	H38g545					yes	FL
697	ORnP	H38g546						
698	OR4AnP	H38g547						
699	OR4An	H38g548					yes	FL
700	OR6Sn	H38g549					yes	FL
701	OR4RnP	H38g550						
702	OR13Cn	H38g551					yes	FL
703	OR13DnP	H38g552						FL
704	OR7EnP	H38g553						FL
705	OR10PnP	H38g554						FL
706	OR8In	H38g555					yes	FL
707	OR8G1	H38g556			HSTPCR25	+	put	
708	ORnP	H38g557						
709	OR5F1	H38g558			OR11-10		yes	FL
710	OR5FnP	H38g559						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
711	OR6BnP	H38g560						FL
712	OR2D1	H38g561			hg27		put	
713	OR5ASn	H38g562					yes	FL
714	OR5SnP	H38g563						FL
715	OR5AQnP	H38g564						
716	OR6BnP	H38g565						FL
717	OR5JnP	H38g566						FL
718	OR9AnP	H38g567						FL
719	OR5BEnP	H38g568						FL
720	OR9An	H38g569					yes	FL
721	OR8Hn	H38g570					yes	FL
722	OR5BNnP	H38g571						
723	OR8Jn	H38g572					yes	FL
724	OR9NnP	H38g573						
725	OR7EnP	H38g574						FL
726	OR7E9P	H38g575		OST289				FL
727	OR8KnP	H38g576						
728	OR2AnP	H38g577						
729	OR8Kn	H38g578					yes	FL
730	OR7E39P	H38g579		OST931	hg611			

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
731	OR7E27P	H38g580		OST932	hg616			
732	OR2Hn	H38g581					put	
733	OR13CnP	H38g582						FL
734	OR13Cn	H38g583					yes	FL
735	OR2S1P	H38g584		OST611				FL
736	OR2AMnP	H38g585						
737	OR1N1	H38g586		OST933	OR1-26		put	
738	OR2S2	H38g587		OST715			yes	FL
739	OR7E26P	H38g588			OR1-51;OR1-72;OR1-73;OR912-95			
740	OR1F11	H38g589			hg91		put	
741	OR5ACnP	H38g590						FL
742	OR5B10P	H38g591			OR13-34;OR13-64;OR13-67			
743	OR2AnP	H38g592						FL
744	OR1E5	H38g593	DS117;DS143		OR13-66	+	put	
745	OR4Fn	H38g594					yes	FL
746	OR5CnP	H38g595						
747	OR2WnP	H38g596						
748	OR2L2	H38g597			HSHTPCRH07	+	put	
749	OR4H8P	H38g598			OR14-58			
750	OR5D10P	H38g599			OR912-94			

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
751	OR7A12P	H38g600			OR14-11;OR14-59			
752	OR2L1	H38g601			HSHTPCRX02	+	put	
753	OR2F3P	H38g602			OR14-60		put	
754	OR4H10P	H38g603		OST934	OR15-69;OR15-80;OR15-81			
755	OR5H1	H38g604			HSHTPCRX14	+	put	
756	OR2K1	H38g605			HSHTPCRX17	+	put	
757	OR7E11P	H38g606			OR11-2			
758	OR7A3P	H38g607		OST935	OR11-7b			
759	OR6A1	H38g608			OR11-55	+	yes	FL
760	OR5I1	H38g609			OLF1	+	yes	FL
761	OR2H3	H38g610			HUMORLMHC	+	yes	FL
762	OR10J1	H38g611	DS3;DS14		HSHGMP07J	+	yes	FL
763	OR7E3P	H38g612			OR11-9			
764	OR1D6P	H38g613			OR11-13;OR11-22			
765	OR5D10P	H38g614			OR18-17;OR18-42;OR18-43;OR18-44			
766	OR5D5P	H38g615			OR18-79;OR912-47			
767	OR52A1	H38g616			HPFH1OR	+	yes	FL
768	OR2AEn	H38g617					yes	FL
769	OR6LnP	H38g618						FL
770	OR6LnP	H38g619						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
771	OR7MnP	H38g620						
772	OR13Cn	H38g621					yes	FL
773	OR13Cn	H38g622					yes	FL
774	OR2InP	H38g623				+		
775	OR4An	H38g624					yes	FL
776	OR2InP	H38g625				+		
777	OR4AnP	H38g626						FL
778	OR4AnP	H38g627						FL
779	OR8C1P	H38g628			OR11-175			
780	OR4AnP	H38g629						FL
781	OR7E15P	H38g630			OR11-392			
782	OR10A1	H38g632			OR11-403		put	
783	OR2An	H38g633				+	put	
784	OR7EnP	H38g634				+		FL
785	OR7En	H38g635				+	put	
786	OR51A1P	H38g636			HPFH6OR	+		FL
787	OR7E47P	H38g637			HSORBPL41;bpl41-16	+		FL
788	OR5B5P	H38g638			OR3-144;OR912-92			
789	OR1F10	H38g639			OR3-145		put	
790	OR8G2	H38g640			HSTPCR120	+	put	

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
791	OR1Sn	H38g641					yes	FL
792	OR4AnP	H38g642						FL
793	OR4AnP	H38g643						FL
794	OR4AnP	H38g644						FL
795	OR4AnP	H38g645						FL
796	OR4AnP	H38g646						FL
797	OR4AnP	H38g647						FL
798	OR4An	H38g648					yes	FL
799	OR4An	H38g649					yes	FL
800	OR7E42P	H38g650		OST001				
801	OR2M3P	H38g651		OST003				
802	OR4H11P	H38g652			OR4-114;OR4-115;OR4-119			
803	OR7E57P	H38g653		OST007				
804	OR2B1P	H38g654			OR5-40;OR5-41		put	
805	OR7E34P	H38g655		OST011				
806	OR7E56P	H38g656		OST013				
807	OR3AnP	H38g657						
808	OR4H5P	H38g658			OR5-39;OR5-84			
809	OR1En	H38g659	DS47;DS115;DS120;DS121;DS123;D			+	put	

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
			S125					
810	OR51CnP	H38g660						
811	OR2WnP	H38g661						FL
812	OR51B1P	H38g662			AF149710			FL
813	OR7E81P	H38g663		OST021				
814	OR7E44P	H38g664		OST022				
815	OR5B7P	H38g665			OR6-55;OR6-57			
816	OR7E36P	H38g666		OST024				
817	OR2A5	H38g667			OR7-138;OR7-141		put	
818	OR5B1P	H38g668		OST936	OR8-122;OR8-123			
819	OR8B8	H38g669			HSTPCR85	+	yes	FL
820	OR8B4P	H38g670			AC002556-D		yes	FL
821	ORnP	H38g671						FL
822	OR8B3	H38g672			AC002556-B		yes	FL
823	OR2Bn	H38g673					yes	FL
824	OR8B6P	H38g674			AC002556-G			FL
825	OR8B5P	H38g675			AC002556-A			FL
826	OR4E2	H38g676			AE000658-A		yes	FL
827	OR8B7P	H38g677			AC002556-F			FL
828	OR11JnP	H38g678						FL
829	OR4E1P	H38g679			AE000658			FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
830	OR10DnP	H38g680						
831	ORnP	H38g681						
832	OR8D2	H38g682			AC002556-E		yes	FL
833	OR11InP	H38g683						FL
834	OR11JnP	H38g684						FL
835	OR10AnP	H38g685	DS12;DS65			+		FL
836	OR8C3P	H38g686			OR912-106;OR912-45;pDJ9j14			FL
837	OR2DnP	H38g687						FL
838	OR4PnP	H38g688						
839	OR7E21P	H38g689		OST035	OR4DG			
840	OR2M1	H38g690		OST037			put	
841	OR7AnP	H38g691						
842	OR5D11P	H38g692			OR8-125;OR8-127			
843	OR7E50P	H38g693			OR8-126			
844	OR7E45P	H38g694		OST049				
845	OR7E77P	H38g695		OST060				
846	OR8B2	H38g696			AC002556-C		yes	FL
847	OR8D1	H38g697		OST004	pDJ9j14		yes	FL
848	OR8B1P	H38g698		OST937	OR11-561			FL
849	OR7A1P	H38g699		OST938	OLF4p;OR19-3;hg513			FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
850	OR7E8P	H38g700			OR11-11a;pDJ392a17			FL
851	OR4DnP	H38g701						FL
852	OR7E80P	H38g702		OST939	pDJ392a17			FL
853	OR4DnP	H38g703						FL
854	OR7E10P	H38g704			AC000385-A			FL
855	OR10B1P	H38g705			AC003956-A;OR19-19			FL
856	OR2InP	H38g706				+		
857	OR4Dn	H38g707					yes	FL
858	OR5ACn	H38g708					put	
859	OR2I1	H38g709			AC004179-A;dJ271M21.7;hs6M1-14	+		
860	OR10H1	H38g710			AC004510	+	yes	FL
861	OR7E59P	H38g711		OST119				
862	OR7E28P	H38g712		OST128				
863	OR5B3	H38g713		OST129			put	
864	OR2A6	H38g714		OST182			put	
865	OR6Cn	H38g715					put	
866	OR7E54P	H38g716		OST185				
867	OR7E48P	H38g717		OST193				
868	OR67AnP	H38g718						FL
869	OR4DnP	H38g719						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
870	OR4CnP	H38g720						FL
871	OR4DnP	H38g721						FL
872	OR10H2	H38g722			AC004597-A	+	yes	FL
873	OR10H3	H38g723			AC004597-B	+	yes	FL
874	OR55CnP	H38g724						
875	OR55BnP	H38g725						
876	OR52VnP	H38g726						FL
877	OR2B3	H38g727			OR6-4;dJ80I19.1;hs6M1-1		yes	FL
878	OR52TnP	H38g728						FL
879	OR2J1P	H38g729			OR6-5;dJ80I19.2;hs6M1-4			FL
880	OR52HnP	H38g730						FL
881	OR2J3	H38g731			OR6-6;dJ80I19.7;hs6M1-3		yes	FL
882	OR52An	H38g732				+	put	
883	OR4Qn	H38g733					put	
884	OR52BnP	H38g734						FL
885	OR2N1P	H38g735	DS9		OR6-7;dJ80I19.3;hs6M1-2	+		FL
886	OR51EnP	H38g736				+		
887	OR2J2	H38g737			OR6-8;dJ80I19.4;hs6M1-6		yes	FL
888	OR2In	H38g738				+	put	
889	OR2J4P	H38g739			OR6-9;dJ80I19.5;hs6M1-5			FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
890	OR7E40P	H38g740		OST215				
891	OR2H4P	H38g741			OR6-3;dJ80I19.6;hs6M1-7			FL
892	OR7E52P	H38g742		OST245				
893	OR2InP	H38g743				+		
894	OR6C1	H38g744		OST267			put	
895	OR7E30P	H38g745		OST339				
896	OR5BAnP	H38g746	DS132			+		
897	OR7H1P	H38g747		OST940	CIT-B-440L2			FL
898	OR5B2	H38g748		OST073			yes	FL
899	OR5AZnP	H38g749						FL
900	OR5Bn	H38g750					yes	FL
901	OR52Bn	H38g751					yes	FL
902	OR5BnP	H38g752						FL
903	OR52Dn	H38g753					yes	FL
904	OR7A11	H38g754		OST527	CIT-HSP-87m17			FL
905	OR5BnP	H38g755						FL
906	OR51AnP	H38g756						FL
907	OR7A15P	H38g757		OST941	CIT-HSP-87m17;OR19-1;OR19-134;OR19-146			FL
908	OR7C2	H38g758			CIT-HSP-87m17;OR19-18		yes	FL
909	OR7E23P	H38g759		OST942	OR21-3			FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
910	OR2E1	H38g760			HS29K1;HSNH0569I24;hs6M1-9			
911	OR1I1	H38g761			F20569;OR19-20		yes	FL
912	OR1RnP	H38g762						FL
913	OR4F3	H38g763			AC004908		yes	FL
914	OR2AEn	H38g764					yes	FL
915	OR2InP	H38g765				+		
916	OR52AnP	H38g766				+		
917	OR7C1	H38g767		OST943	CIT-HSP-146e8;OR19-5;TPCR86	+	yes	FL
918	OR2A3P	H38g768			AC004889-B			FL
919	OR7A5	H38g769	DS8;DS19;DS61;DS68;DS112	OST944	HTPCR2	+	yes	FL
920	OR2InP	H38g770	DS72			+		
921	OR7A10	H38g771		OST027	CIT-HSP-146e8		yes	FL
922	OR2An	H38g772				+	put	
923	OR2M2	H38g773		OST423			put	
924	OR7A8P	H38g774		OST042	OR19-11;hg83			FL
925	OR2An	H38g775				+	put	
926	OR7E20P	H38g776		OST516				
927	OR2AnP	H38g777				+		
928	OR5BHnP	H38g778				+		
929	OR1En	H38g77					put	

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
		9						
930	OR1EnP	H38g780						
931	OR5Bn	H38g781					yes	FL
932	OR8RnP	H38g782						
933	OR5ANn	H38g783					yes	FL
934	OR5ANnP	H38g784						FL
935	OR5BRnP	H38g785						FL
936	OR2A1	H38g786			AC004889-A	+	yes	FL
937	OR10An	H38g787					yes	FL
938	OR2A9	H38g788	DS149		HSDJ0798C17	+		FL
939	OR2A7	H38g789			HSDJ0798C17	+	yes	FL
940	OR10A3	H38g790			HSHTPCRX12	+	yes	FL
941	OR10Cn	H38g791					yes	FL
942	OR7A2P	H38g792			OLF4p;OR19-18;hg1003		yes	FL
943	OR10WnP	H38g793						FL
944	OR7A17	H38g794			HSHTPCRX19		yes	FL
945	OR5Bn	H38g795					yes	FL
946	OR5BnP	H38g796						FL
947	OR1Q1	H38g797		OST226	HSTPCR106;OR9-A;hRPK-465_F_21	+	yes	FL
948	OR2Hn	H38g798	DS133;DS144;DS150			+	yes	FL
949	OR7EnP	H38g79						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
		9						
950	OR7A14	H38g800		OST945	OR19-12			
951	OR1B1	H38g801			OR9-B;hRPK-465_F_21		yes	FL
952	OR12D2	H38g802			AC004171;dJ994E9.8;hs6M1-20	+	yes	FL
953	OR7EnP	H38g803						FL
954	OR8BnP	H38g804						FL
955	OR1L1	H38g805			OR9-C;hRPK-465_F_21;hg23		yes	FL
956	OR11An	H38g806					yes	FL
957	OR7AnP	H38g807						
958	OR1C1	H38g808			HSTPCR27	+	yes	FL
959	OR1D2	H38g809		OST946	OR17-4	+	yes	FL
960	OR1L3	H38g810			OR9-D;hRPK-465_F_21		yes	FL
961	OR12DnP	H38g811						FL
962	OR4G1P	H38g812			OLB			FL
963	OR2B4P	H38g813			AL050339-A;dJ974I11.1;hs6M1-22			
964	OR11H1	H38g814			OR22-1		yes	FL
965	OR4Fn	H38g815					yes	FL
966	OR56AnP	H38g816						FL
967	OR8NnP	H38g817						FL
968	OR7EnP	H38g818						
969	OR4Pn	H38g81					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
		9						
970	OR6Cn	H38g820					put	
971	OR5BCnP	H38g821						
972	OR10QnP	H38g822	DS64			+		FL
973	OR5BnP	H38g823						FL
974	OR10PnP	H38g824						FL
975	OR1L4	H38g825		OST046	OR9-E;hRPK-465_F_21		yes	FL
976	OR2APnP	H38g826						
977	OR1L6	H38g827		OST947	HShRPK-465_F_21;hg16		yes	FL
978	OR6UnP	H38g828						FL
979	OR5C1	H38g829			OR9-F;hRPK-465_F_21		yes	FL
980	OR11InP	H38g830						FL
981	OR4AnP	H38g831						FL
982	OR4GnP	H38g832						FL
983	OR10Vn	H38g833					yes	FL
984	OR4G2P	H38g834			HS14a-1-B			FL
985	OR10VnP	H38g835				+		
986	OR4F4	H38g836			HS14a-1-A		yes	FL
987	OR4G3P	H38g837			OLC-7501			FL
988	OR5AKnP	H38g838						FL
989	OR10YnP	H38g839						FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
990	OR4GnP	H38g840						FL
991	ORnP	H38g841						
992	OR4Fn	H38g842					yes	FL
993	OR8A1	H38g843		OST025			yes	FL
994	OR8Bn	H38g844					yes	FL
995	OR6DnP	H38g845						
996	OR7E14P	H38g846		OST948	OR11-5	+		FL
997	OR2M4	H38g847		OST710	HSHTPCRX18	+	put	
998	OR4WnP	H38g848						
999	OR4Fn	H38g849	DS36			+	yes	FL
1000	OR7EnP	H38g850						
1001	OR4GnP	H38g851						FL
1002	OR10JnP	H38g852						
1003	OR52En	H38g853					yes	FL
1004	OR4RnP	H38g854						FL
1005	OR4Cn	H38g855					yes	FL
1006	OR4AnP	H38g856						
1007	OR4AnP	H38g857	DS54			+		
1008	OR4AnP	H38g858						FL
1009	OR9Gn	H38g859					yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
1010	OR10An	H38g860					yes	FL
1011	OR4Cn	H38g861					yes	FL
1012	OR10VnP	H38g862						
1013	OR10UnP	H38g863						FL
1014	OR7E2P	H38g864	DS127		OR11-6;hg94	+		FL
1015	OR7E35P	H38g865		OST018				FL
1016	OR9KnP	H38g866						
1017	OR7E13P	H38g867		OST949	OR11-4			FL
1018	OR7EnP	H38g868						FL
1019	OR9Kn	H38g869					yes	FL
1020	ORnP	H38g870						FL
1021	OR7EnP	H38g871		OST950	OR11-1;hg500	+		FL
1022	OR7EnP	H38g872						FL
1023	OR3A4P	H38g873		OST951	OR17-24;OR17-25	+	yes	FL
1024	OR8QnP	H38g874						
1025	OR7EnP	H38g875						FL
1026	OR7EnP	H38g876						FL
1027	OR3A1	H38g877	DS2		OLFRA03;OR17-40;hg138	+	yes	FL
1028	OR5Gn	H38g878					yes	FL
1029	OR5MnP	H38g879						

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
1030	OR7EnP	H38g880						FL
1031	OR5G1P	H38g881		OST952	OR11-104;OR93;OR93Hum			FL
1032	OR5PnP	H38g882						FL
1033	OR10AEnP	H38g883						
1034	OR3A2	H38g884		OST953	OR17-228	+	yes	FL
1035	OR10Jn	H38g885					yes	FL
1036	OR1D3P	H38g886		OST954	OR17-23			FL
1037	OR10Jn	H38g887					yes	FL
1038	OR1D4	H38g888			OR17-30	+	yes	FL
1039	OR5GnP	H38g889						FL
1040	OR4SnP	H38g890						FL
1041	OR5GnP	H38g891						FL
1042	OR9HnP	H38g892						FL
1043	OR1A1	H38g893			OR17-7	+	yes	FL
1044	OR1A2	H38g894			OR17-6	+	yes	FL
1045	OR8AnP	H38g895						FL
1046	OR1P1P	H38g896			OR17-208	+		FL
1047	OR7E12P	H38g897		OST955	AC000378-A;OR11-3;hg1058	+		FL
1048	OR4A1P	H38g898			OR11-30			FL
1049	OR10G3	H38g899			AE000658-D		yes	FL

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
1050	OR10G1P	H38g900			AE000658-C			FL
1051	OR10G2	H38g901			AE000658-B		yes	FL
1052	OR5Tn	H38g902					yes	FL
1053	OR7EnP	H38g903						FL
1054	OR7EnP	H38g904						FL
1055	OR4AnP	H38g905						FL
1056	OR4C1	H38g906			HSHTPCRX11	+		FL
1057	OR1EnP	H38g907						
1058	OR7KnP	H38g908						FL
1059	OR4CnP	H38g909						FL
1060	OR1RnP	H38g910						FL
1061	OR5AUn	H38g911					yes	FL
1062	OR4Cn	H38g912					yes	FL
1063	OR4Cn	H38g913					yes	FL
1064	OR13DnP	H38g914						FL
1065	OR5n	H38g915	DSU116			+		
1066	OR2Hn	H38g916	DSU150			+		
1067	ORn	H38g917	DSU151			+	put	
1068	ORn	H38g918	DSU17			+		
1069	ORn	H38g919	DSU18			+		

SEQ ID #	Symbol	HORDE	Digi	OST	Trivial	Tran	Int.	E
1070	ORn	H38g920	DSU35			+		
1071	OR6Fn	H38g921	DSU41			+		
1072	ORn	H38g922	DSU49			+		
1073	ORn	H38g923	DSU50			+		
1074	OR10An	H38g924	DSU57			+		
1075	ORn	H38g925	DSU58			+		
1076	OR2Ln	H38g926	DSU59			+		
1077	OR10Jn	H38g927	DSU60			+		
1078	OR1Kn	H38g928	DSU63			+		
1079	OR10Dn	H38g929	DSU7			+		
1080	ORn	H38g930	DSU32			+		
1081	OR2Ln	H38g931	DSU38			+		
1082	ORn	H38g932	DSU62			+		
1083	ORn	H38g933	DSU48			+		
1084	OR2n	H38g934	DSU111			+		

Table 2

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
153	OR10D3	0	11	137.96	...........SDVISV	69	M	AC074177.4	12106 ... 13038
154	OR7EnP	4	4	11.58	MVACGVLDLHIIDSFAL	53	R	AF091580.1	7 ... 663
155	OR1D5	0	17	3.75	LVVTNLLYLLLLTGIFT	49	M	AF073967.1	2 ... 649

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
156	OR10NnP	4	11	138.02	LQGSGVVHILFGNVLAT	82	M	AC074177.4	159287 ... 158526
157	OR2F1	0	7	148.62	LLGGFTSSVQIISSLLT	56	M	AF073974.1	41 ... 649
158	OR7EnP	7	4	11.58	MAGGELLDLHILPALGL	54	M	AF073989.1	547 ... 1515
159	OR8FnP	6	11	137.96	LLVICEMGAHCVCSNIF	75	M	AC069561.10	51687 ... 50743
160	OR2Q1P	2	7	148.62	LLCGFSANMEIVSGVIL	49	M	AC020865.3	190954 ... 189954
161	OR2W1	0	6	33.74	LMGSCMINVLLVLGIVT	88	M	AF102516.1	52 ... 669
162	OR7EnP	7	4	11.58	MVACGVLDLHITHSFGL	53	R	AF091580.1	7 ... 663
163	OR6B1	0	7	148.62	LIMCCGIIAKFDLAIFF	61	M	NM_010983.1	178 ... 975
164	OR10Kn	0	1	154.34	MLGSSACVVTLILGALI	79	M	AC073778.1	168744 ... 167803
165	ORnP	13	11	138.02	VPYCIGGHLLICLSLSS	33	M	AC074177.4	12106 ... 13038
166	OR4F2P	4	6	186.49	IHGGMVLHFQFVNSICG	50	M	AB030896.1	1 ... 906
167	OR7EnP	3	4	11.58	MVACGVLDLHIIDSFGL	54	M	AF102536.1	22 ... 669
168	OR1F2P	0	16	6.15	MSADNGVNLHLIEAVTT	72	R	M64377.1	1 ... 939
169	OR2P1P	7	6	33.74	FGGSCMSNQSALVRXSV	48	M	NM_008762.1	1 ... 936
170	OR7E43P	5	4	5.57	MAGGELFDLHIMPAFGL	54	M	AF102536.1	22 ... 669
171	OR4F1	4	6	0.23	IHGGMVLHFQFVNSICG	50	M	AB030896.1	1 ... 906
172	OR7E55P	5	3	89.94	MAGDEFLDLHILPAFGL	53	M	AF073989.1	547 ... 1515
173	OR13Dn	0	9	86.89	MLGSCWITLQLMTNSLI	61	M	AC023789.5	371264 ... 372220
174	OR4CnP	3	16		AHGAIVGHIQFVNSICL	74	M	AF102522.1	40 ... 660
175	OR10D1P	1	11	137.96	LHGCCGFQFLLGSVMPS	83	M	AC074177.4	128803 ... 129726
176	OR4Cn	0	16		LHGGIVGHVQLVNSICL	86	M	AB030895.1	1 ... 924

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
177	OR8GnP	0	11	137.96	LSAICGLGIHFVLSNIM	73	M	AC074177.4	106297 ... 105361
178	OR13CnP	2	9	86.85	MFGACGGNLQLMASFLG	82	M	AJ251154.1	2703 ... 1747
179	OR4CnP	5	16		LHEAIVLHIQFINSLCL	61	M	AF102522.1	40 ... 660
180	OR13Cn	0	9	86.81	MLGTCGINVQFMATFIT	69	M	AJ133425.1	61 ... 1014
181	OR4CnP	0	16		LHGGIMGHIQLVNSMCL	63	M	AB030895.1	1 ... 924
182	OR51Bn	0	11		AHSVSGRSPVRPLITIL	76	M	AF071080.2	15931 ... 16851
183	OR7E5P	2	11	51.76	MVACDVLDLHIIDSFGL	54	M	AF073989.1	547 ... 1515
184	OR13Cn	0	9	86.77	MFGSCVSNVQLMSNFLL	71	M	AJ251154.1	2703 ... 1747
185	OR4Sn	0	16		LHGGIAAHLQLVNSISA	56	M	AB030895.1	1 ... 924
186	OR51BnP	4	11		VHYPEWRSPPPPLVIFL	72	M	AF071080.2	15931 ... 16851
187	OR6JnP	1	14	2.72	CFGTFFGSFPLDLSVIC	50	R	M64378.1	1 ... 933
188	OR51Bn	0	11		SHAISGRSPISPQTTVL	76	M	AF071080.2	26330 ... 27262
189	OR7EnP	2	11	71.8	MFACGVLDLHIIDSFGL	55	M	AF102536.1	22 ... 669
190	OR2An	0	6	144.32	TSAVCTTLIHLVGAGLG	81	M	L14566.1	62 ... 667
191	OR7E22P	3	3	89.94	MVACDVLDLHIIDSFGL	56	M	AF073989.1	547 ... 1515
192	OR7E4P	2	11	71.8	IVACDVLDLHIMHSFGL	55	M	AF102536.1	22 ... 669
193	OR7E66P	9	3	89.94	MAGGELLFLHIMPAFGL	55	M	AF073989.1	547 ... 1515
194	OR6Mn	0	11	138.18	TFGTFGGSFPVNLSVIS	50	M	NM_010991.1	1 ... 939
195	OR2ALnP	11	11	112.69	ILGTCASNFDFFNHLLL	32	M	AL359352.1	85325 ... 86251
196	OR6MnP	2	11	138.18	TGGTFGGSCPVNLSILT	50	M	NM_010991.1	1 ... 939
197	OR4D1	0	17	60.7	IHGGVAGHVQLMNSLVI	90	M	AC019272.4	62255 ... 61317
198	OR5D2P	3	11	51.09	LCVVTTWCTLFTSANES	48	M	AC073947.3	29192 ... 30115

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
199	OR7E38P	7	7	95.91	MAGGELFHLHIMPAFGL	55	R	AF091580.1	7 ... 663
200	OR4D2	0	17	60.7	IHGGVAGHVQLKNSLDV	89	M	AC019272.4	183633 ... 182701
201	OR7E7P	4	7	95.91	MIACGVLDLHIIDSFGL	56	R	AF091580.1	7 ... 663
202	OR5AHnP	0	19	68.97	...........RSGIMC	77	M	AC020957.2	48184 ... 49107
203	OR2U2P	5	6	33.53	LVYSCIVNIPYTMCIVV	49	M	AC044846.2	105668 ... 104736
204	OR2U1P	2	6	33.53	LVCTCMINILCCVVIFA	54	M	AF102516.1	52 ... 669
205	OR2H2	0	6	33.19	ILGTCVIEVQSVASILV	89	M	AL078630.1	41097 ... 40165
206	OR2H5P	7	6	33.19	FLGTCVIEVQSMASILV	84	M	AL078630.1	41097 ... 40165
207	OR2In	0	6	33.19	LLGSCASNAQLMARILL	74	M	AL078630.1	151152 ... 150391
208	OR11HnP	5	13		IFNTCLCWIPLCLSVIG	60	M	AF121972.1	171 ... 1109
209	OR7EnP	6			AAACDVIDLHITHSFGL	56	M	AF073964.1	41 ... 649
210	OR9In	0	11	54.06	FTAGCGCGLRCIFGVIA	50	R	AF091579.1	7 ... 663
211	OR2AFnP	11	X	140.17	MLGTCGHVTLAGISTLL	43	R	L34074.1	73 ... 1011
212	OR13KnP	5	X	140.17	MFGMCVIIIHLGIGTLL	43	R	L34074.1	73 ... 1011
213	OR13Cn	0	9	86.77	MFGSCVSNVQLLSNFLL	68	M	AJ251154.1	2703 ... 1747
214	OR13Fn	0	9	86.77	MLGSCGTTVESMISLLM	55	M	AJ133428.1	61 ... 1017
215	OR9Qn	0	11	54.08	FTGSCGASVRSIFAVIA	47	M	AF146372.1	509 ... 1456
216	OR2TnP	1	1	254.77	ILIGFGGDMLVMCCMLI	71	M	AF102527.1	22 ... 669
217	OR4Kn	0	14	0.08	IHVGMIVHSHFTNSISS	56	M	AF259072.1	104176 ... 105099
218	OR2B8P	0	6	31.6	LLGSCTINLQLLVSILV	62	R	L34074.1	73 ... 1011

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
219	OR2Tn	0	1	254.77	MLAGVALDLLITCCMLT	57	M	AF102527.1	22 ... 669
220	OR4Kn	0	14	0.08	IHTGIAMHSQFMTSIAS	53	M	AF259072.1	104176 ... 105099
221	OR2A4	0	6	144.76	TSAVCTTLIHLVGAGLG	81	M	L14566.1	62 ... 667
222	OR7EnP	6	2	161.53	MVACDVLDLHIIDSFGL	54	R	AF091580.1	7 ... 663
223	OR4Kn	0	14	0.08	MHGGILVHSQFMTSIAV	57	M	AF259072.1	104176 ... 105099
224	OR13InP	6	9	86.85	MYGSCVLNNVVIGKTLL	41	M	AJ251155.1	15491 ... 16423
225	OR7EnP	8	2	161.53	MVACDVLDLHIFFDFGL	54	M	AF073989.1	547 ... 1515
226	OR6Jn	0	14	2.72	CFGTFFGSFPLDLSVIC	50	R	M64378.1	1 ... 933
227	OR4Mn	0	14	0.08	LHGAMLGHIQLMSSISV	54	M	AC019272.4	183633 ... 182701
228	OR4VnP	10	11	51.09	IHGIIVLHFQMVNSFAV	50	M	AB030896.1	1 ... 906
229	OR6Xn	0	11	138.36	AFGTFSVICQLGATVIG	46	M	AF106007.1	178 ... 975
230	OR51Gn	0	11	3.7	LHSSSSRLPLLGVVTVV	55	M	NM_013617.1	1 ... 921
231	OR6EnP	3	14	2.72	SFGTFCTLIPLGIASLG	82	M	NM_010991.1	1 ... 939
232	OR4NnP	2	14	0.08	LHGGGAGHIQLMNSMTL	54	M	AC019272.4	62255 ... 61317
233	OR6MnP	7	11	138.18	IFGTFGGARLVSXSMVT	37	R	M64378.1	1 ... 933
234	OR4Nn	0	14	0.08	LHGGGAGHIQLMNSMTL	57	M	AC019272.4	62255 ... 61317
235	OR4Cn	0	11	51.09	LHGGIGGHIQFVNSMCA	65	M	AF102522.1	40 ... 660
236	OR4KnP	4	14	0.08	IHAGMGTHSQFMDSMGT	51	M	AF259072.1	104176 ... 105099
237	ORnP	8	11	137.59	AIAITVVVAHAAAGVVA	35	M	AC069559.8	73704 ... 74636
238	OR5D3	0	11	51.15	FCVVTAWCTYFISANES	46	R	U50948.1	34 ... 978
239	OR2G1P	6	6	33.53	LLGSCVSNIQVLASLLL	84	M	AL359352.1	85325 ... 86251

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
240	OR4Kn	0	14	0.08	IHTGMIVHSQFINSLSS	51	M	AF259072.1	104176 ... 105099
241	OR8BnP	2	11	137.59	LCVFSGMGAHNVIVGIV	68	M	AC069559.8	120212 ... 119283
242	OR2B2	0	6	31.47	LLGSCASNLQWLISFLI	89	R	L34074.1	73 ... 1011
243	OR7EnP	3	2	73.87	MVACDVLDLRIIDSFGL	54	M	AF073989.1	547 ... 1515
244	OR4KnP	3	14	0.08	IHTGIVVHSQFMTSIAI	57	M	AB030896.1	1 ... 906
245	OR2AD1P	6	6	33.87	FLGACTSSIVLVFGFLV	51	M	AL136158.14	162423 ... 161461
246	OR1AAnP	8	X	140.17	MIVDNTIVLHLIIGVII	48	M	AC068902.11	144125 ... 143193
247	OR1E3P	1	17	2.99	MLGVSLLHLHLMMGILI	74	R	M64392.1	1 ... 942
248	OR8BnP	3	11	137.59	FCVFSGMGAHNIVVGIV	63	M	AC069561.10	96653 ... 95690
249	OR5Hn	0	3	104.18	FAGTCFGHIHLVLSIQF	55	R	AF091575.1	52 ... 663
250	OR1G1	0	17	2.99	LMVMAAMHLHLITGTGI	56	R	M64392.1	1 ... 942
251	OR5HnP	2	3	104.18	FAVTCGGHIHFVFSIQF	46	M	AC068904.15	165039 ... 165965
252	ORnP	5	X	140.17	MLVTCSHHFLSFTGIWS	36	R	U50948.1	34 ... 978
253	ORnP	11	X	140.17	LIVTFAKITTTQDHHHH	29	M	AC069561.10	127636 ... 126698
254	OR4PnP	2	11	51.09	LHGDIAGHSQLVNSISL	51	M	AB030895.1	1 ... 924
255	OR13Hn	0	X	140.17	TLATCTTVAMLITSTLL	47	M	AJ251154.1	35662 ... 36615
256	OR7D1P	5	19	11.38	VMAGTAIFVHLLATLGF	64	R	AF091580.1	7 ... 663
257	OR4KnP	2	18	47.77	IHNGIVVHSQFMTSIAI	55	M	AB030896.1	1 ... 906
258	OR7E24	1	19	11.38	MVACDLIDLHIIMGFGL	60	R	AF091580.1	7 ... 663
259	OR51NnP	2	11	3.6	LHGFSARSPSLGVLVTV	49	R	AF079864.1	632 ... 1576
260	OR7E18P	6	19	11.38	VAGCDLLDLHIMLAFGL	59	M	AF102536.1	22 ... 669

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
261	OR7E19P	2	19	11.38	MYVCDVLNLHIMDSFGL	58	M	AF073989.1	547 ... 1515
262	OR7E41P	7	11	14.36	IVVCDMLDLHIHSTFGL	55	M	AF073989.1	547 ... 1515
263	OR2R1	3	7	148.69	LLGGFVVNMELISSVLV	77	M	AF073974.1	41 ... 649
264	OR10ACnP	7	7	148.69	MVGGCGRVGLLLACLLL	46	M	AC073778.1	168744 ... 167803
265	OR51Ln	0	11	3.79	LHTFSARVPTLGVVTLV	54	R	AF079864.1	632 ... 1576
266	OR52JnP	3	11	3.79	MHTGSSRLPILGVALDA	57	M	AF121979.1	53 ... 1106
267	OR9LnP	9	8	45.22	TVVNNFFFFFFIFDLIA	37	M	AC069561.10	147203 ... 146274
268	OR51PnP	4	11	3.79	MHSISARLPALGVVSML	48	M	AF071080.2	2641 ... 1697
269	OR5HnP	4	3	104.18	FAVTCLGHIHFFFSIQL	50	R	AF091575.1	52 ... 663
270	OR51An	0	11	3.79	EHSVSVKLPFTYFGCLV	48	R	AF079864.1	632 ... 1576
271	OR5HnP	6	3	104.18	FAVTCLGHIHFVFSIQF	46	M	AC068904.15	165039 ... 165965
272	ORnP	11	17	17.43	LLPCILSIIALYYYYYY	27	M	AL359352.1	9138 ... 8177
273	OR52En	0	11	3.79	MHTGSARFPFFYCAILF	57	M	AF121979.1	53 ... 1106
274	OR5Hn	0	3	104.18	FVVTCLGHIHFVFAVQF	53	R	AF091575.1	52 ... 663
275	OR4CnP	3	11	50.21	VHRGVVGHIQFVNSICL	73	M	AF102522.1	40 ... 660
276	OR52En	0	11	3.79	MHTLSGRFPSLYCANLF	60	M	AF121979.1	53 ... 1106
277	OR10Dn	0	11	138	LHGCCGIHILLGNVLSI	86	M	AC074177.4	12106 ... 13038
278	OR5HnP	2	3	104.18	FVVTCLGHIHFVFAIQF	54	R	AF091575.1	52 ... 663
279	OR13An	0	10	47.91	LTASLALNIHLIADYGV	67	M	AF102520.1	16 ... 669
280	OR5HnP	2	3	104.18	FGGTCLGHIHILLSIQF	57	R	AF091575.1	52 ... 663

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
281	OR5Kn	0	3	104.47	FCETCGAHIHLLFSVQF	45	M	AC069559.8	36251 ... 35322
282	OR7EnP	9	21	17.99	MAGGELFHLQIMPAFGL	57	M	AF073989.1	547 ... 1515
283	OR4DnP	6	8	77.48	IHGGVAGHVQVMNSLVI	87	M	AC019272.4	62255 ... 61317
284	OR2ARnP	0	3	30.89	MLGSC............	71	M	AJ251154.1	56533 ... 57369
285	OR7E29P	4	3	136.03	MAGGELLDLHIMPAFGL	56	M	AF073989.1	547 ... 1515
286	OR4CnP	3	11	51.12	AHGAIVGHIQFVNSICL	74	M	AF102522.1	40 ... 660
287	OR5PnP	2	11	6.93	LVGTCVGNTFCPSSIIV	74	M	AF121977.1	262 ... 1197
288	OR7EnP	5	3	136.04	MVACGVLDLHIIGSFGL	52	R	AF091580.1	7 ... 663
289	OR56An	0	11	4.73	MNLPSFRLPILQAGLLS	41	M	AF121975.1	50 ... 1012
290	OR56AnP	9	11	4.73	KNQAFFRMPILQGGLLS	73	M	AF121981.1	89 ... 475
291	OR5Pn	0	11	6.89	LAATCVAISYSLSSIIV	63	M	AF121977.1	262 ... 1197
292	OR7E53P	5	3	136.04	MAGGEFPDLHIMPAFGL	54	M	AF073989.1	547 ... 1515
293	OR5Pn	0	11	6.89	LVGTCMGNTFCPSSIIA	83	M	AF121977.1	262 ... 1197
294	OR52Ln	0	11	4.73	MHSSSVRLPFLGMAVIL	59	M	AF121976.2	474 ... 1307
295	OR5E1	3	11	6.89	LGATXGYNIQLLFSNLG	51	R	U50948.1	34 ... 978
296	OR56AnP	3	11	4.73	MNLASFRMAILPPPPPP	39	M	AF121976.2	474 ... 1307
297	OR4KnP	2	8	88.25	IHTGMIVHSQFIDS...	57	M	AB030896.1	1 ... 906
298	OR52Ln	0	11	4.73	MHSSSVRLPFLGVAVVL	59	M	AF121976.2	474 ... 1307
299	OR7EnP	1	4	74.82	MVF..............	55	R	AF091580.1	7 ... 663
300	OR52XnP	5	11	4.73	MHSASLXLSFLAVALGG	51	M	AF121976.2	474 ... 1307
301	ORnP	13	4	74.82	STGCKGRKXLKLVRDFQ	24	R	M64386.1	130 ... 975
302	OR56An	0	11	4.73	MNLTSFRVPVLQAGLLS	84	M	AF121981.1	89 ... 475

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
303	OR56AnP	10	11	4.73	LI...GMMXNL...KKK	60	M	AF121981.1	89 ... 475
304	OR1R1P	5	17	3	MVGISAVHLHLIEGVVA	48	M	AF073967.1	2 ... 649
305	OR52EnP	2	11	3.79	MHTGSGRSPFLYGAILF	64	M	AF121979.1	53 ... 1106
306	OR51AnP	4	11	3.7	EHTVALKLPLLGAGSTL	46	R	AF079864.1	632 ... 1576
307	OR51An	0	11	3.7	EHSVSVKLPFTYFGCLV	48	R	AF079864.1	632 ... 1576
308	OR4CnP	1	11	51.12	VHGGVVGHVQFVNSICL	75	M	AF102522.1	40 ... 660
309	OR52JnP	9	11	3.79	MHTGACRFPILGVVYLN	58	M	AF121979.1	53 ... 1106
310	OR4RnP	9	11	51.12	.....GGGVXSVNGNYL	66	M	AF102522.1	40 ... 660
311	OR52Jn	0	11	3.79	MHTGACRLPMLGVVFVN	58	M	AF121976.2	474 ... 1307
312	OR4CnP	3	11	51.12	VHGGGVGHIQFINSICL	76	M	AF102522.1	40 ... 660
313	OR51AnP	2	11	3.79	EHSASAKLPFTYFVTGL	83	M	AF121985.1	2 ... 478
314	OR7EnP	15	12	93.55	IVVCDLLDLHIHSTFGL	55	M	AF073989.1	547 ... 1515
315	OR5MnP	2	11	52.17	CIVLHVYLMERMVASNQ	54	M	AF102528.1	52 ... 669
316	OR10ABnP	1	11	6.93	MLASCAVFCITILSVLG	47	M	AC073778.1	168744 ... 167803
317	OR52SnP	2	11	3.79	MHSTSARLPHLSVATGV	54	M	AF121976.2	474 ... 1307
318	OR5Mn	0	11	52.14	CIVHIFYTAAWMLANFY	49	R	AF091579.1	7 ... 663
319	OR10Sn	0	11	138.1	LHASCIIHIHLMSIVAG	61	M	AF259072.1	32953 ... 32000
320	OR5MnP	4	11	52.14	CIVHIFYTTAWMLANFY	48	R	AF091579.1	7 ... 663
321	OR10Gn	0	11	138.1	LHGSCGSHVQLIDIVAG	61	M	AF259072.1	55611 ... 54658
322	ORnP	20	11	29.15	ILGIYEGSAHYFIILFL	33	M	AL365337.1	192661 ... 191711
323	OR5MnP	2	11	52.19	CIVIYGYSMEWMVANLS	54	M	AF102528.1	52 ... 669

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
324	OR10GnP	10	11	138.1	LYGSCWGHLPIYVIKFT	30	M	L14567.1	17 ... 667
325	OR10TnP	1	1	154.34	LVACCACTIVLILSVLV	57	M	X92969.1	8035 ... 8961
326	ORnP	16	11	52.17	LAAPLLLVFVLAAAAAA	33	R	M64376.1	1 ... 999
327	OR10RnP	11	1	154.5	MLAVFTICVFLIGGALV	47	M	AC023611.2	108224 ... 107271
328	OR5MnP	2	11	52.16	CIVHLVYTMEWMVANFY	49	R.	AF091579.1	7 ... 663
329	OR7EnP	4	8	6.68	MLACGVLDLHIIDSFGL	55	M	AF102536.1	22 ... 669
330	OR10Tn	0	1	154.27	LLACCLTIVALLLSVIV	58	M	AC012302.5	54283 ... 55224
331	OR1E1	0	17	3.04	MLGDSLLHLHLIMGILI	83	R	Y07557.1	1 ... 942
332	OR5BKnP	4	12	42.11	STGGAIAIMDFLSQWGL	46	M	AF073965.1	2 ... 643
333	OR5MnP	3	11	52.17	CIVHIVYTMEWMVANLF	48	R	AF091579.1	7 ... 663
334	OR3A3	0	17	3.06	LHAGCACNTHALAAMAA	49	M	AF073967.1	2 ... 649
335	OR10ADnP	1	12	42.11	TFGVCTFNFLIIDAVIS	44	M	AF247657.1	1 ... 945
336	OR10Rn	0	1	154.5	MLAICAGATVLICGVLV	56	M	AC073778.1	168744 ... 167803
337	OR5TnP	4	11	51.94	MCGTCAAHIHAFFVIEV	51	M	AF121977.1	262 ... 1197
338	OR4GnP	15	7	0.23	ICRKMAVHSQFVNSISA	42	M	AB030892.1	1 ... 939
339	OR6Yn	0	1	154.5	LVVCYGCTIKFDLAVII	61	M	NM_010983.1	178 ... 975
340	OR1E2	0	17	3.15	MLSDSLLHLHLIMGILI	80	R	Y07557.1	1 ... 942
341	OR8Hn	0	11	51.94	MVGACGINVNWILATLV	51	M	NM_013728.1	1 ... 948
342	OR4Fn	0	7	0.23	IHGGMVIHSQFVNSLTC	50	M	AC019272.4	62255 ... 61317
343	OR10Kn	0	1	154.27	MLGCSACVIILILCVLI	83	M	AC073778.1	168744 ... 167803
344	OR7LnP	11	X	140.17	MLGVCGHGTNLXFFFFI	32	M	AL133160.1	63932 ... 64759
345	OR8InP	7	11	51.94	MVVCCMINVSVSLATLG	44	R	M64386.1	130 ... 975

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
346	OR10RnP	0	1	154.5	MLAVCTSIVGFIFGVLV	54	M	AC073778.1	168744 ... 167803
347	OR2AFnP	11	X	140.17	MLGTCGHVTLAGISTLL	43	R	L34074.1	73 ... 1011
348	OR8Kn	0	11	51.94	LEIILVYVFLKIFSNLF	55	M	AF102528.1	52 ... 669
349	ORnP	7	10	127.57	S.CCCLLTYIIHHHHHH	31	M	AC020958.1	164590 ... 163746
350	OR8KnP	10	11	51.94	MIIILIYQMVKIFSNLF	35	M	AC073945.4	152209 ... 153150
351	OR51Hn	0	11	3.6	MHGISSRVPVLGVVTLL	49	R	AF079864.1	632 ... 1576
352	OR7EnP	5	3	136.03	MVACGVLDLHIIDSFGL	51	M	AF073989.1	547 ... 1515
353	ORnP	8	3	56.17	LLLLFLIIEQH.....I	32	R	M64376.1	1 ... 999
354	OR5BMnP	20	3	103.93	KXNKCTLSSSLMVFIQF	30	M	AF146372.1	509 ... 1456
355	OR10GnP	0	11	138.1	LHGCCGGHFQFTDILAT	63	M	AF259072.1	55611 ... 54658
356	OR2Yn	0	5	209.23	LLGSCAANIQLMARVVV	74	M	AC044846.2	139468 ... 138536
357	OR10DnP	1	11	138.1	LHGCCGGHVLLSNVVAM	66	M	AC074177.4	128803 ... 129726
358	OR3BnP	7	X	158.48	IHAPSILNTYLLSFVAA	37	M	AL136158.14	29455 ... 30402
359	OR8Dn	0	11	138.1	LCVICAVDIHCIIGNMA	62	R	X80671.1	203 ... 1129
360	OR5RnP	0	11	52.13	LLMICVYVFHIIFADMS	68	M	AF102528.1	52 ... 669
361	OR10Gn	0	11	138.1	LHGSCGSHVQLINIVAG	58	M	AF259072.1	55611 ... 54658
362	OR5BDnP	12	11	53.74	MTGTCVVIHRALSSITP	39	M	NM_013728.1	1 ... 948
363	OR5ALnP	1	11	52.13	VIVVLSYVVQALIANTC	52	M	AC073947.3	29192 ... 30115
364	OR52HnP	3	11	4.15	LHFVSGRVPCLGVPTVT	59	M	AF121975.1	50 ... 1012
365	OR10Gn	0	11	138.1	LHGGCSSHVQLITVVAG	56	M	AF259072.1	55611 ... 54658

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
366	OR5Mn	0	11	52.17	CIVHIVYTMEWMVANLF	52	M	AF146372.1	509 ... 1456
367	OR51Mn	0	11	4.15	MHSFSIRAPILGVVTVL	50	M	NM_013617.1	1 ... 921
368	OR6Tn	0	11	138.1	SFGTFAAWCPLALSVLG	52	M	NM_010991.1	1 ... 939
369	OR6DnP	5	10		SLGSFVVLGLKALVVLT	69	R	AF034903.1	85 ... 1053
370	OR4B1	0	11	45.36	IHGVIGGHIQVVNSFSF	62	M	AF102522.1	40 ... 660
371	OR5ALnP	4	11	52.13	VISVVGYMIQALIANVC	50	M	AF146372.1	509 ... 1456
372	OR51Qn	0	11	4.15	FHSFSACAPSLGLAIIV	49	M	NM_013617.1	1 ... 921
373	OR4Dn	0	11	138.1	LHGGIAGHVQLMNNVTM	63	M	AC019272.4	62255 ... 61317
374	OR52Nn	0	11	4.58	MHTGSLRLPSLGVAIGF	52	M	NM_013619.1	118 ... 969
375	OR4Xn	0	11	45.36	MHGGAIGHGQLINGISV	58	M	AB030896.1	1 ... 906
376	OR8Jn	0	11	52.03	LLIVVLYTVVYVSANVG	77	M	X89682.1	2 ... 472
377	OR51JnP	2	11	4.15	MHSMSIKLPLLGIVTFL	46	M	AF071080.2	15931 ... 16851
378	OR10Gn	0	11	138.1	LHGSCSSHVQLIDIVAG	60	M	AF259072.1	55611 ... 54658
379	OR52En	0	11	4.58	MHTGTVRLPFLGVIIID	66	M	AF121979.1	53 ... 1106
380	OR4Xn	0	11	45.36	LHGGIIGHAQLINGLSI	64	M	AB030895.1	1 ... 924
381	OR10A2	1	11	5.69	MFGVCAPVVQWAGTVVI	76	M	AF247657.1	1 ... 945
382	OR5Mn	0	11	52.14	CIVHVVYVICWMIANFY	49	R	AF091579.1	7 ... 663
383	OR52En	0	11	4.58	MHTGSVRFPFLISVVGI	59	M	AF121979.1	53 ... 1106
384	OR8Kn	0	11	51.94	LLIGLIYILVKIFADLS	53	M	AF146372.1	509 ... 1456
385	OR10An	0	11	5.66	MFGACASVVQWAATFIF	89	M	AF247657.1	1 ... 945
386	OR8LnP	3	11	52.13	LIVVMSYVLQLLLANTF	51	M	AF102528.1	52 ... 669
387	OR5BPnP	8	11	52.82	VVVVVGGSIVPPVGLHL	43	R	U50948.1	34 ... 978
388	OR52Nn	0	11	4.58	MHTGSARLPFLGVAIGF	54	M	AF121976.2	474 ... 1307

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
389	ORnP	7	11	45.36	WWWWWIALLR.AAAAAK	28	M	X89686.1	32 ... 472
390	OR8JnP	1	11	51.94	LLIVILQTTVCVFSNLF	99	M	X89682.1	2 ... 472
391	OR5Mn	0	11	52.24	CIVIFVYNSQLMVATLS	50	R	AF091579.1	7 ... 663
392	OR52En	0	11	4.58	MHTVSIRMPLLGSILLL	66	M	AF121979.1	53 ... 1106
393	OR5Tn	0	11	51.94	VCGTCAAHIHALFVIEV	52	M	AF146372.1	509 ... 1456
394	OR52NnP	5	11	4.58	MHTGSVQLPFLGAAIGF	51	M	NM_013619.1	118 ... 969
395	OR4B2P	6	11	45.36	IFGIIGRHVQVVNSELS	53	M	AB030896.1	1 ... 906
396	OR51KnP	6	11	4.15	MHSCSGKLPLLGIVNFL	51	M	NM_013617.1	1 ... 921
397	OR52QnP	10	11	4.58	MYTGSVRFPFLFVAVGI	45	M	AF121979.1	53 ... 1106
398	OR4Fn	0	15	86.21	IHGGMIIHIQFVNSISA	50	M	AF102522.1	40 ... 660
399	OR11MnP	1	12	41.92	FSAACGSSFTL......	48	M	AL359381.1	175785 ... 176720
400	OR52Nn	0	11	4.44	MHTGSARLPFLGVAIGF	57	M	NM_013619.1	118 ... 969
401	OR56An	0	11	4.58	MNLASFRMPILQGGLLS	73	M	AF121981.1	89 ... 475
402	OR5AWnP	14	X		LXADFTSNLPTTSSNVV	39	R	X80671.1	203 ... 1129
403	OR52Nn	0	11	4.51	MHTGSARLPFLGVAIGF	55	M	AF121976.2	474 ... 1307
404	ORnP	15	X		ISCIFELTLPLPSNVNV	31	M	AC073947.3	29192 ... 30115
405	OR52EnP	6	11	4.58	VHSVSVRMPILGNIILL	62	M	AF121979.1	53 ... 1106
406	OR5BHnP	9	X		MVASCGGKTVSLCGTLT	40	M	NM_013728.1	1 ... 948
407	OR4QnP	1	15	1.66	IHGAMAGHMQLMNSLSV	60	M	AC019272.4	62255 ... 61317
408	OR51En	0	11	3.04	MHSGSARLPLFGVIAIL	60	R	AF079864.1	632 ... 1576
409	OR11KnP	2	15	1.66	FSGYGFCITLLITFVFI	53	M	AF121972.1	171 ... 1109
410	OR12D1P	1	6	33.02	LHGSATIHLHMSTGIAG	76	M	AL133159.4	16108 ... 15185

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
411	OR4NnP	3	15	1.61	LHGGGAGHIQLMNSMTM	55	M	AC019272.4	62255 ... 61317
412	OR11A1	0	6	33.02	FGATCTSVLVLTLSCLI	76	M	AL359381.1	175785 ... 176720
413	OR10C1	0	6	33.02	MLGACSCVGHFIATLIC	59	M	AL365336.1	122764 ... 121784
414	OR2H1	0	6	33.02	LLGTCVMQVQSLSSFVV	88	M	AL078630.1	48786 ... 47851
415	OR9RnP	8	12	59.71	LAVGGGCNIQFLLSITT	54	R	AF091579.1	7 ... 663
416	OR4FnP	0	7	0.53	.....VLHFQFVNSICG	50	M	AB030896.1	1 ... 906
417	OR7D4	3	19	11.31	VMAGTAIFVHLLATLGF	67	R	AF091580.1	7 ... 663
418	OR7E25P	3	19	11.31	MIACSVLDLHIVIGFGL	61	R	AF091580.1	7 ... 663
419	OR2D2	0	11	5.69	LLGCCGSVVDFITGILI	65	M	AF073987.1	2 ... 649
420	OR10An	0	11	5.69	MFGVCAPVVQWAGTVVI	76	M	AF247657.1	1 ... 945
421	OR2WnP	3	1	254.49	LLGGCVCQGHWVLAVVS	54	R	L34074.1	73 ... 1011
422	OR7E16P	8	19	11.31	IAGCDLLDLHIMLALGL	60	M	AF102536.1	22 ... 669
423	OR52Pn	0	11	4.44	MHCMSARLPCLGAAVIV	59	M	AF121976.2	474 ... 1307
424	OR6AnP	4	11	5.66	LLGCCGGIVKLDLAILG	94	R	M64386.1	130 ... 975
425	OR7D2	0	19	11.24	VMPITVITLHLIMTLGF	61	R	AF091580.1	7 ... 663
426	OR52UnP	3	11	4.44	LHSASVRFPMLGVAVAY	52	M	AF121976.2	474 ... 1307
427	OR2AGn	0	11	5.6	MLGGDTLSIYYVMGFLP	55	M	AF102527.1	22 ... 669
428	OR7G3	0	19	11.24	ILVGNLVDLHMVVTLGV	64	R	AF091580.1	7 ... 663
429	OR56BnP	3	11	4.44	IHVGSFRFPVLQLAGMS	41	M	AF133300.1	25713 ... 26573
430	OR2AGnP	1	11	5.51	MLGSDTLIGHYITGFLL	55	M	AF102527.1	22 ... 669
431	OR56Bn	0	11	4.44	MHVASFRCSVLQLALMS	39	M	NM_013619.1	118 ... 969
432	OR6AnP	5	11	5.51	LLGCCGGIVKLDLAILG	93	R	M64386.1	130 ... 975
433	OR4FnP	4	19	63.23	IHGGMVLHFQFVNSICG	49	M	AB030896.1	1 ... 906

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
434	OR6Wn	0	7	148.04	SFGSFAVSSPQDLSFVT	47	M	NM_010991.1	1 ... 939
435	OR4Mn	0	15	1.59	LHGAMLGHIQLMSSISV	52	M	AF259072.1	104176 ... 105099
436	OR52YnP	13	11	3.6	VVVVVLQWPVMGMAVDF	29	M	AF133300.1	46551 ... 47498
437	OR11HnP	2	15	1.78	FFGTCLCWIPLCLSVIG	61	M	AF121972.1	171 ... 1109
438	OR9An	0	7	148.04	LSGTFVFSWPALMAILG	46	M	NM_010991.1	1 ... 939
439	OR5Mn	0	11	52.19	CILLFFYDFQLMSANLS	50	M	AC069563.9	129775 ... 130725
440	OR6Vn	0	7	148.04	FFGSFAAAPTSDMAFVS	45	M	NM_010991.1	1 ... 939
441	OR4Nn	0	15	1.61	LHGGGAGHIQLMNSMTL	53	M	AC019272.4	62255 ... 61317
442	OR51AnP	4	11	3.6	EHTDSLILPFTGLACMS	43	M	NM_013617.1	1 ... 921
443	OR9PnP	10	7	148.04	FGSNSFEHLVFIHSLLM	39	M	NM_010983.1	178 ... 975
444	OR4H6P	3	15	1.66	MHGCILGHVQLVNSISG	59	M	AF259072.1	104176 ... 105099
445	OR51FnP	2	11	3.6	MHTFSLRLPLLGDLTTI	48	R	AF079864.1	632 ... 1576
446	OR7E1P	3	11	68.1	MVACGVLDLHIIDSFGL	55	M	AF073989.1	547 ... 1515
447	OR51Tn	0	11	3.6	MHSLSVRFPLAGLQLNT	44	R	AF079864.1	632 ... 1576
448	OR2Vn	0	13	104.15	IVVGGSFDIQVICCMLF	84	M	AF102535.1	16 ... 669
449	OR51HnP	7	11	3.6	MHGGSARAPVLGAVIIL	51	R	AF079864.1	632 ... 1576
450	OR51An	0	11	3.6	EHTVSIRLPFTGIACTL	48	M	AF071080.2	26330 ... 27262
451	OR2AInP	2	5	209.13	YLGSCLSNFHLMARILL	55	M	AC044846.2	112743 ... 113748
452	OR2F2	0	7	148.74	LLGGFTSNVQIISSLLT	54	M	AF073974.1	41 ... 649
453	OR1F12	0	6	31.61	MMANNAINLHMVTVIFV	58	M	AC023167.7	60743 ... 61663

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
454	OR7G1P	0	19	11.24	ILAGSLMDVQMIASFGI	60	R	AF091580.1	7 ... 663
455	OR7G2	0	19	11.24	ILAGNLTNLLMIAAFGV	61	R	AF091580.1	7 ... 663
456	OR1M1	0	19	11.24	MHGISAFITHLIVAVIT	89	M	X89689.1	32 ... 472
457	OR51UnP	1	11	2.89	VTDDN...........	48	R	AF079864.1	632 ... 1576
458	OR52Hn	0	11	4.19	MHFVSGRIPDLGVPTVS	59	M	AF121975.1	50 ... 1012
459	OR1F1	0	16	6.15	MFVDNGVNLHLIEGVMT	75	R	M64377.1	1 ... 939
460	OR10PnP	0	16	87.09	MIGICTTTTHLVATFII	48	M	AF247657.1	1 ... 945
461	OR4FnP	4	19	7.9	IHGGMVLHFQFVNSICG	49	M	AB030896.1	1 ... 906
462	OR2T1	0	1	254.77	HLVGFGGDLLIMCCMLI	92	M	AF102527.1	22 ... 669
463	OR7EnP	9	19	22.8	VAGCDLLDLHIMLAFGL	60	M	AF102536.1	22 ... 669
464	OR51Gn	0	11	3.6	LHSFSVRLPLMGVITVI	57	M	NM_013617.1	1 ... 921
465	OR2Tn	0	1	254.77	MVAGFGLDTFIMCCMLI	67	M	AF102527.1	22 ... 669
466	OR5BGnP	2	11	51.27	AAAAAGGSIHNLFAVEI	52	R	U50948.1	34 ... 978
467	OR5WnP	3	11	51.27	MGADCLVDIHCMFVVAC	51	M	AF146372.1	509 ... 1456
468	OR51Sn	0	11	3.6	MHSVSARLPLLLVLMGD	42	M	AF071080.2	26330 ... 27262
469	OR5WnP	1	11	51.27	...........LVFIES	55	M	AC074177.4	107189 ... 107708
470	OR51AnP	3	11	3.6	EHTDSLILLPTGVAMMD	46	M	NM_013617.1	1 ... 921
471	OR5Dn	0	11	51.21	FCGVTGWCILFCIANES	46	M	AF146372.1	509 ... 1456
472	OR7EnP	4	4	5.55	MVACGVLDLHIIDSFGL	54	R	AF091580.1	7 ... 663
473	OR51Fn	0	11	3.6	MHTFSSRVPVFGALTTF	53	R	AF079864.1	632 ... 1576
474	OR5Dn	0	11	51.21	YCVVSGWGVLYLFANEC	48	M	NM_013728.1	1 ... 948
475	OR52Rn	0	11	3.6	VHSSSIRWPFMGVAVAF	58	M	AF121976.2	474 ... 1307
476	ORnP	27	11	51.21	FCFAAGQSPGFLCFFFF	23	M	AB030893.1	37 ... 930

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
477	OR7EnP	6	3	121.47	MVACDVLDLHIIDSFSL	57	M	AF073989.1	547 ... 1515
478	OR6Qn	0	11	54.04	LTGACAVTLPLDVSVLA	52	M	NM_010983.1	178 ... 975
479	OR4Fn	0	6	185.89	IHGGMVLHFQFVNSICG	51	M	AB030896.1	1 ... 906
480	OR7EnP	3	13	40.31	FFSP.AAALHIMPAFGL	65	M	X89686.1	32 ... 472
481	OR7En	0	2	95.17	MVACDVLDLHIIDSFGL	57	M	AF073989.1	547 ... 1515
482	OR4Nn	0	14	0.27	LHGAMVGHVQLMNSLSL	58	M	AC019272.4	62255 ... 61317
483	OR2ASnP	7	1	254.77	.....GGGGGMICGLLP	43	M	AF102535.1	16 ... 669
484	OR11Hn	0	14	0.33	FFGTCFIGIPYFQSVLF	90	M	AF121972.1	171 ... 1109
485	OR2Tn	0	1	254.77	MLAGFGLDMLIMCCMLI	69	M	AF102527.1	22 ... 669
486	OR2TnP	1	1	254.77	CMMGFSGDLLIMCCMLI	77	M	AF102527.1	22 ... 669
487	OR2AKnP	3	1	254.55	TLGGACSNIHYVSGILL	50	M	AF102533.1	16 ... 669
488	ORnP	16	12	4.38	VLKSKCWQLPFYMPLLM	25	R	Y07557.1	1 ... 942
489	OR5DnP	4	11	51.21	FCAVTGWSTLFCIANES	48	R	U50948.1	34 ... 978
490	OR7EnP	1	4	5.55	FVACDVLDLHIIDNFGL	54	M	AF102536.1	22 ... 669
491	OR5L2	0	11	51.27	FCGVVCCCIHLLVANEV	53	M	AF146372.1	509 ... 1456
492	OR5Dn	0	11	51.27	FCVVLVWCTLSLVANES	48	M	NM_013728.1	1 ... 948
493	ORnP	4	9	81.99	..CCCLFFQSIASGTYI	23	M	AL359381.1	82137 ... 81544
494	OR10Qn	0	11	54.08	MVGSCGLPQLLLVSVLI	50	M	AL365336.1	123248 ... 124093
495	OR9MnP	1	11	51.27	LCVDSGGSIHNLFAVEI	54	M	AC069559.8	73704 ... 74636
496	OR7E62P	5	2	73.96	MAACDVLDLHTIDSFRL	56	M	AF073989.1	547 ... 1515
497	OR9LnP	13	11	54.06	MFVGCTLVAYGILTMIA	32	M	AC069561.10	147203 ... 146274

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
498	OR7E46P	10	2	73.96	MAGVEFCDLHIMPAFGL	54	M	AF102536.1	22 ... 669
499	OR1S1	0	11	54.08	MIVVNILITHLLVGVIF	56	M	AC073769.1	133488 ... 132556
500	OR5DnP	0	11	51.21	FCVIMGWCTLSCISSEC	45	M	AC069563.9	111696 ... 112671
501	OR9InP	4	11	54.06	FTASCGGNICCISAVIT	46	R	AF091579.1	7 ... 663
502	OR5Dn	0	11	51.21	FCVVSGWCELSLLANES	53	M	AF146372.1	509 ... 1456
503	OR9QnP	4	11	54.08	FTASCGASVRTIFAVMA	47	M	AL365337.1	192661 ... 191711
504	OR51CnP	0	11	3.04	MKTVSARMPMLGAMTVV	51	R	AF079864.1	632 ... 1576
505	OR5WnP	1	11	51.27	FCADCGVDIHL......	53	M	AC069561.10	127636 ... 126698
506	OR9InP	2	11	54.06	FTAGCSCGLHCICAMFA	46	M	AC074177.4	106297 ... 105361
507	OR51AnP	4	11	3.04	MHSVSARVPVPGVVTGL	72	M	X89685.1	2 ... 481
508	OR5L1	0	11	51.21	FCVVVCCCIHLLVANEV	55	M	AF146372.1	509 ... 1456
509	OR7EnP	5	13	50.42	.....VVDLHIMPAFGL	66	M	X89686.1	32 ... 472
510	OR5BLnP	18	11	54.08	ILGNXLENQCFIFAMIT	29	R	M64392.1	1 ... 942
511	OR51En	0	11	3.04	MHSASVRFPLLGAIVMV	95	R	AF079864.1	632 ... 1576
512	OR51Dn	0	11	3.04	MHSASSRFPLIGIIVMV	61	R	AF079864.1	632 ... 1576
513	OR52In	0	11	3.04	MHTATARFPLMSGSMVS	46	M	AF121975.1	50 ... 1012
514	OR4KnP	2	18	19.04	IHTGMIVHSQFIDSLSS	56	M	AB030896.1	1 ... 906
515	OR52In	0	11	2.99	MHTATARAPLMSGSMVS	47	M	AF121975.1	50 ... 1012
516	OR4KnP	2	18	19.04	IHNGIVVHSQFMTSIAI	55	M	AB030896.1	1 ... 906
517	OR52MnP	1	11	3.04	MHATSVRYLPIGIGVLL	51	R	AF079864.1	632 ... 1576

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
518	ORnP	7	6	31.58	FLVSCLLLLLLLEGIHW	30	M	AF073964.1	41 ... 649
519	ORnP	9	8	88.25	IXVVVLNIVNMTTIIFL	24	M	AC074177.4	149899 ... 148964
520	ORnP	9	10	70.63	YSIVMFYHAHFICELLN	26	M	AC068902.11	144125 ... 143193
521	ORnP	9	9	70.7	WWWWWSWYGNFDDSITX	26	R	AF091563.1	7 ... 669
522	ORnP	9	5	202.43	FFFFF.PPPPP......	27	R	AF034902.1	4197 ... 5177
523	ORnP	10	11	137.77	LLLLWSQFXQFLAVVVV	29	R	M64376.1	1 ... 999
524	ORnP	3	11	16.31	NNNNNLLXMNILTLLAI	27	M	AL136158.14	29455 ... 30402
525	ORnP	17	11	55.6	LAGNNIYCYHM..LLLL	26	R	M64377.1	1 ... 939
526	OR6Pn	0	1	154.6	LIACCASSMKFDLAMIL	60	M	NM_010983.1	178 ... 975
527	OR7EnP	3	14	33.48	MVACDVLDLHIIDSFGL	54	R	AF091580.1	7 ... 663
528	ORnP	12	11	138.51	LMCHS.FFFFFMMMMMM	29	R	AF091573.1	7 ... 663
529	OR7EnP	5	14	33.48	MAGGDFLDLYILPDFGL	55	M	AF073989.1	547 ... 1515
530	ORnP	7	10	127.4	S.CCCLLTYIIHHHHHH	31	M	AC020958.1	164590 ... 163746
531	OR10XnP	2	1	154.6	MLGGCSAITELIISGLG	49	M	AC073778.1	168744 ... 167803
532	OR10Zn	0	1	154.71	MAACCTTFGMVILSVLV	56	M	AC025913.3	108128 ... 109067
533	OR6KnP	2	1	154.73	MYGIVGCTPEWVVHEIT	40	R	M64386.1	130 ... 975
534	OR6Kn	0	1	154.73	MHGIVSCTPEWVIHEIT	44	M	AC027184.3	54955 ... 54017
535	OR1FnP	1	4	97.57	...........IEGVMT	73	R	M64377.1	1 ... 939
536	OR1ABnP	3	19	19.44	MIGISAFNTHLV.....	64	M	AC073769.1	133488 ... 132556
537	OR52MnP	1	11	2.89	MHATSARYLPIGIGVLL	49	M	AF121975.1	50 ... 1012
538	OR1XnP	6	5	202.43	MIANTLGIVHIFAALFA	71	M	AF102530.1	1 ... 666

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
539	OR4FnP	8	16	83.04	QQQQQVIHSQFVNSLTC	46	M	AC019272.4	62255 ... 61317
540	OR52MnP	5	11	2.89	MHATSVRYLPIGIGVLM	45	R	AF079864.1	632 ... 1576
541	OR2Vn	0	5	209.61	IVVGGSFDIQVICCMLF	83	M	AF102535.1	16 ... 669
542	OR2V1P	4	5	209.61	IVVGGSFDIQALCCMLL	90	M	AF102537.1	16 ... 669
543	OR2Zn	0	19	65.55	ITGVGSVNIQILSGILL	76	M	AC073769.1	54319 ... 55289
544	OR52KnP	5	11	2.89	...........AMFIEL	52	M	AF121975.1	50 ... 1012
545	OR10Hn	0	19	19.7	MFGFSWGMMVIGLVTAI	75	M	AC023604.2	214343 ... 213396
546	OR2Dn	0	11	5.77	ILGCCRSVVDFIMGILA	85	M	AF073987.1	2 ... 649
547	OR7EnP	6	2	161.49	VVGGCSSDLHIMPAFGL	64	M	X89686.1	32 ... 472
548	OR11GnP	4	14	0.27	FFGSCSLWIPVSLSLLI	68	M	AC027184.3	54955 ... 54017
549	ORnP	12	14	0.27	GSCGNSLHHYLMVNIIL	28	M	AF121972.1	171 ... 1109
550	OR11Gn	0	14	0.33	FFGSCNLWIPNFLSPVM	67	M	AF121972.1	171 ... 1109
551	OR11HnP	5	14	0.33	FTGTAFFSVSQFLSIIL	68	M	AF121972.1	171 ... 1109
552	OR6Kn	0	1	154.73	MHENGGFIPEMDHATII	46	R	AF034897.1	354 ... 1199
553	OR11Hn	0	14	0.33	FFGTCVGCVPLCFNIIG	71	M	AF121972.1	171 ... 1109
554	OR6KnP	0	1	154.73	MHGNGGFVPEWDHAAIF	46	M	AL365336.1	122764 ... 121784
555	OR11HnP	2	14	0.33	FFGTCLIGISFFVSFIL	70	M	AF121972.1	171 ... 1109
556	OR6KnP	2	1	154.82	MHGVAGFMPECDRASIT	43	M	AC027184.3	54955 ... 54017
557	OR6Kn	0	1	154.84	MHGISGCLPEWVIHEIA	45	R	AF034900.1	1 ... 963
558	OR2Ln	0	1	254.55	SSGGAGINAHYVSTFLF	53	M	AF102527.1	22 ... 669
559	OR4GnP	8	16	83.04	ICRKMAVHSQFVNSISA	45	M	AB030892.1	1 ... 939

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
560	OR6Nn	0	1	154.84	IHGACGGGVELDINKIA	50	R	M64386.1	130 ... 975
561	OR2LnP	2	1	254.55	SLAVGGINAHY.....W	52	M	AF102535.1	16 ... 669
562	OR9A1	0	7	146.91	LLGTLVLSWPALMAIIG	45	M	L14567.1	17 ... 667
563	OR6Nn	0	1	155.69	THGACACCSELDINIII	51	M	AL136158.14	29455 ... 30402
564	OR10Hn	0	19		MFGFSCGMVVAGLVTAL	86	M	AC023604.2	245345 ... 246298
565	OR7EnP	4	9	71.72	MVACDVLDLHIMNSFGL	57	M	AF073989.1	547 ... 1515
566	OR2AQnP	5	1	155.69	FCHSCLLLLSLLPFFFF	31	M	AL359352.1	55588 ... 56546
567	OR2LnP	3	1	254.55	SMAGAGINAHYVSSFLF	50	M	AF102537.1	16 ... 669
568	OR5ARn	0	11	52.46	FVVDCGASAHLLLCIES	53	R	AF091579.1	7 ... 663
569	OR7EnP	4	9	71.79	TAGGETLDLHIMPAFGL	57	M	AF102536.1	22 ... 669
570	OR10AAnP	2	1	155.69	THGMCAAAVPLHVIATC	84	M	AC005992.15	9114 ... 8173
571	OR10JnP	4	1	157.7	MIAICGVVVQSNVSVIV	72	M	X92969.1	8035 ... 8961
572	OR5A1P	0	11	55.81	FVGLCGGSIQSNVVVGT	81	M	Y15525.1	1 ... 705
573	OR2AHnP	5	11	52.46	MLGSCISSVILVFSIVI	51	M	AF247657.1	1 ... 945
574	OR10JnP	4	1	157.7	LLGICGIMVQSNVSVLL	68	M	X92969.1	8035 ... 8961
575	OR56BnP	2	11	4.93	IHMCSSRLPVLQLVVVS	39	M	AF121975.1	50 ... 1012
576	OR5M1	0	11	52.35	CIVIFIYSSQLMVANLS	49	R	AF091579.1	7 ... 663
577	OR52WnP	0	11	4.93	MHTASLLAVPLGLSISM	48	M	AF121976.2	474 ... 1307
578	OR5AMnP	5	11	52.35	FIVIYAYNVQLMVANLC	35	M	AC068904.15	113793 ... 114719
579	OR52BnP	3	11	4.93	MHFVSTQTPVLGVPSVV	89	M	AF121975.1	50 ... 1012
580	OR5MnP	1	11	52.35	CVLLYFWVMQLLSANLV	48	R	X80671.1	203 ... 1129

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
581	OR5APnP	6	11	52.35	FGAGGALNIHFIFANES	55	R	X80671.1	203 ... 1129
582	OR56Bn	0	11	4.95	IHFCSFRLPVLQLALVS	41	M	AF121975.1	50 ... 1012
583	OR5APn	0	11	52.35	FGLGCTANIHMIFSIVS	55	M	AF121977.1	262 ... 1197
584	OR52Bn	0	11	4.93	GHFVSARIPVLGVPMVL	73	M	AF121975.1	50 ... 1012
585	OR9Gn	0	11	52.5	FAAYCVGNIIKMLLNVC	45	M	AC074177.4	106297 ... 105361
586	OR52Kn	0	11	2.86	MHSISARLPLLGVASVL	53	M	NM_013619.1	118 ... 969
587	OR5MnP	1	11	52.35	FIVIYAYNSQLMVANLC	51	M	AC074177.4	106297 ... 105361
588	OR52Kn	0	11	2.86	MHSISARLPLLGVAIVL	52	M	NM_013619.1	118 ... 969
589	OR52KnP	3	11	2.82	MHSISARLPLLGVAIGL	53	M	NM_013619.1	118 ... 969
590	OR52BnP	4	11	2.78	IHFISARVPDLGVLTVL	57	M	AF121975.1	50 ... 1012
591	OR2B6P	0	6	31.62	LLGAYATNWLLLVSFHI	79	R	L34074.1	73 ... 1011
592	OR2WnP	7	6	31.61	LLRGCASNVMLAFAIVL	58	M	AF102516.1	52 ... 669
593	OR2AnP	5	7	148.83	TMAHCTCLVHLISSILG	72	M	AF102521.1	22 ... 669
594	ORnP	16	6	31.61	FLVSCMDFMYIVLNNVI	39	M	AF102516.1	52 ... 669
595	OR2LnP	0	1	254.55	STAVAGINAHYVSAFLF	50	M	AF102527.1	22 ... 669
596	OR2W2P	5	6	31.61	LLGGCVCQSYWVLSIVM	55	R	L34074.1	73 ... 1011
597	OR2LnP	1	1	254.55	SLAGA............	61	M	AF102535.1	16 ... 669
598	OR2B7P	1	6	31.61	LLGGCTTNIQLIVSFLV	59	M	AC044846.2	105668 ... 104736
599	OR2Ln	0	1	254.43	SLGGAGINAHYVSAFLF	53	M	AF102527.1	22 ... 669
600	OR5BFn	0	1	254.77	VVVYLASYMHSISAVGG	46	M	AL359352.1	9138 ... 8177

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
601	OR2LnP	4	1	254.55	SVAGMSMDAHYVSTFLF	47	M	AF102527.1	22 ... 669
602	OR7EnP	3	10	17.14	MVACCVLDLHI......	51	R	AF091580.1	7 ... 663
603	OR1H1	2	9	106.04	LGADNVIHVHLLVALLA	57	M	AC073769.1	133488 ... 132556
604	ORnP	14	1	254.49	TTTKKSERIYIVSSFLI	24	M	AF102527.1	22 ... 669
605	OR4Dn	0	11	55.81	IHGGIASHIQLMNNVTL	64	M	AC019272.4	183633 ... 182701
606	OR1Ln	0	9	106.04	MYGNSFFHLHLQEAVLT	54	M	AC023167.7	60743 ... 61663
607	OR5AXn	0	1	254.2	LTSAIVIFAYGGVGLSS	47	M	AL136158.14	154973 ... 155908
608	OR5An	0	11	55.77	YCGLCGGSIESTVSVGV	64	M	Y15525.1	1 ... 705
609	OR5AYn	0	1	254.2	LVAGILNLLYGSIGYAS	50	M	AL359352.1	126933 ... 127889
610	OR13Gn	0	1	255.42	LTLGMMINVHLVADLAG	59	M	AF102540.1	16 ... 669
611	OR5BBnP	0	11	55.77	YASLCGGSVHPLEAVGG	54	M	Y15525.1	1 ... 705
612	OR9GnP	6	11	52.49	FVXNCAGNIIELMLNIT	47	M	AF121977.1	262 ... 1197
613	OR2TnP	4	1	254.77	HLAGFAGNLLVMCCMLI	75	M	AF102527.1	22 ... 669
614	ORnP	7	1	255.42	PVAGKGAFLHSVESLGS	38	M	AL365337.1	192661 ... 191711
615	OR1Jn	0	9	95.9	MITDSVLSSHLMVGVIL	66	M	AF102524.1	52 ... 669
616	OR2CnP	1	16	6.47	LLGACIGNIQFLVCFTV	85	M	M84005.1	1 ... 936
617	OR9GnP	2	11	52.49	FAAYCYGNILNLLLNVS	49	M	AL365337.1	192661 ... 191711
618	OR2C1	0	16	6.4	LLGACIGNIQFLVCFTV	85	M	M84005.1	1 ... 936
619	OR51AnP	2	11	4.22	....	52	M	AF071080.2	26330 ... 27262
620	OR9Gn	0	11	52.49	LCAYCGGNAHNLVVTVS	53	M	AC068904.15	165039 ... 165965

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
621	OR52Bn	0	11	2.78	LHFISTRTPILGILTVL	61	M	AF121975.1	50 ... 1012
622	OR1K1	0	9	105.89	MFGVSMVHLYLIEGVVT	58	R	M64377.1	1 ... 939
623	OR51RnP	3	11	2.78	MHTYSARLPGLGSISLL	47	R	AF079864.1	632 ... 1576
624	OR7EnP	2	13	54.83	MVACDVLDLHILDSFGL	57	M	AF073989.1	547 ... 1515
625	OR52PnP	3	11	2.82	MHSASARLPLLGAAVVT	55	M	AF121975.1	50 ... 1012
626	OR7EnP	5	9	70.7	MVACDVQYVHSMDSFGL	48	M	AF102536.1	22 ... 669
627	OR7EnP	5	9	70.7	TAGGD.CCCCC......	43	M	AF073989.1	547 ... 1515
628	OR4KnP	1	21	8.12	IHTGMIVHSQFIDSLSS	57	M	AF259072.1	104176 ... 105099
629	OR4KnP	2	21	8.12	IHNGIVVHSQFMTSTAT	54	M	AB030896.1	1 ... 906
630	OR7EnP	6	9	70.7	.....VFLVHSVPAFGL	58	M	X89686.1	32 ... 472
631	OR51In	0	11	4.15	MHSFSGKTPFVGVITYM	51	R	AF079864.1	632 ... 1576
632	OR51In	0	11	4.15	MHSMSGRTPLLGVLTFM	56	R	AF079864.1	632 ... 1576
633	OR2AnP	1	7	148.83	TLAICTFL.........	63	M	AF102521.1	22 ... 669
634	OR2A2	2	7	148.83	TLAVCTCLVHLITCVLG	68	M	AF102521.1	22 ... 669
635	OR2AnP	8	7	148.83	TFAACTCLVHLITCVLG	68	M	AF102521.1	22 ... 669
636	OR2Gn	0	1	256.63	LHGSCMSTVQLLASFLV	59	M	NM_008762.1	1 ... 936
637	OR2AnP	0	7	148.83	TLAHCAFFFFL......	57	M	AF102521.1	22 ... 669
638	OR6Fn	0	1	254.2	MFGCYGCAVPLAIAVIS	71	R	M64378.1	1 ... 933
639	OR2AnP	4	7	148.83	TLAHCAFLVHLISCILG	68	M	AF102521.1	22 ... 669
640	OR2Gn	0	1	256.02	LLGSCISSIHFLVSFVI	63	M	M84005.1	1 ... 936
641	OR7E37P	5	13	26.5	MAGGEFLDLHIMPAFGL	57	M	AF073989.1	547 ... 1515
642	OR5AVn	0	1	256.02	AMATVMSCMHAVFGLVI	51	M	AL359352.1	9138 ... 8177

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
643	OR2AJnP	7	1	254.43	VLLGCGINVHYVSAFLI	55	M	AF102527.1	22 ... 669
644	OR13EnP	1	9	39.89	MLGSCLTNLQLLATLTA	79	M	AJ251155.1	15491 ... 16423
645	OR2Cn	0	1	257.85	FHGACAGTVGLMASFVL	63	M	M84005.1	1 ... 936
646	OR2TnP	0	1	254.43	IPGGCSLDLQAMCCMLV	59	M	AF102537.1	16 ... 669
647	OR2WnP	2			LMGSCVCNIMQTLGLLV	56	M	M84005.1	1 ... 936
648	OR13Jn	0	9	39.89	MLGSCALKTEILGSLLV	82	M	AJ251155.1	6062 ... 6997
649	OR6RnP	2	1	254.39	SFGCFLGLPSLDSSLIS	45	M	NM_010983.1	178 ... 975
650	OR5ATn	0	1	254.39	VLASLVYIMHGLINLDC	50	M	AL359352.1	111313 ... 112242
651	OR2Zn	0	19	10.64	ITGVGSVNIQILSGILL	76	M	AC073769.1	54319 ... 55289
652	OR4Ln	0	14	0.08	MHGGMLIHSQLVDSLST	53	M	AB030893.1	37 ... 930
653	OR4UnP	14	14	0.15	RHSGMAMHSQLVDSLSL	46	M	AB030895.1	1 ... 924
654	OR4Fn	0	6	185.98	IHGGMIIHIQFVNSISA	50	M	AF102522.1	40 ... 660
655	OR4FnP	2	6	185.98	IHGGMAIHVQFVNSISS	50	M	AB030896.1	1 ... 906
656	OR4Fn	0	6	185.98	IHGGMATHVQFVNSISG	50	M	AB030896.1	1 ... 906
657	OR4Fn	0	6	185.98	IHGGMTIHVQFVNSISG	50	M	AB030896.1	1 ... 906
658	OR4AnP	5	11	50.28	IHGGILGHVQFVNDICV	65	M	AF102522.1	40 ... 660
659	OR4LnP	1	14	0.21	KHGSMLIHSQLVDSLST	53	M	AB030893.1	37 ... 930
660	OR7E33P	6	13	54.79	MAGGEFLDLRILPAFGL	56	M	AF073989.1	547 ... 1515
661	OR2Cn	0	1	257.85	FHGACAGTVGLMASFVL	63	M	M84005.1	1 ... 936
662	OR4Kn	0	14	0.15	MHGGMSVHSQFVDSLSV	53	M	AF259072.1	104176 ... 105099
663	OR5U1	0	6	33.45	VIASVAASMHILFTAAI	84	M	AL359352.1	111313 ... 112242
664	OR4Kn	0	14	0.08	IHGGMAVHSQFMDSLSS	58	M	AF259072.1	104176 ... 105099

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
665	OR5V1	0	6	33.45	LVVGCSANVHLLTGIGT	84	M	AL365337.1	192661 ... 191711
666	OR4QnP	1	14	0.08	LHGAMAGHVQLMNSISI	62	M	AF259072.1	104176 ... 105099
667	OR12D3	0	6	33.45	LHGSAAIYMHMLVTISG	70	M	AL359381.1	128169 ... 127234
668	OR4Kn	0	14	0.08	IHTGMIVHSQFIDSLSS	59	M	AF259072.1	104176 ... 105099
669	OR51CnP	3			MKTVSARMPMLGAMTVV	53	R	AF079864.1	632 ... 1576
670	OR1J2	0	9	105.94	MITDSVLSSHLMVGVIL	66	M	AF102524.1	52 ... 669
671	OR5BJnP	3			SIGSAAVNTKFPSCLGV	46	M	AF073965.1	2 ... 643
672	OR1J1	0	9	105.82	TIADSGICLHLIAAAIL	63	M	AF102524.1	52 ... 669
673	OR13En	0			MLGSCLTNLQLLATLTA	83	M	AJ251155.1	15491 ... 16423
674	OR4KnP	5	14	0.08	IHGGMVIHTHFVNSLSM	53	M	AB030893.1	37 ... 930
675	OR1LnP	5	9	105.84	MYGNSFFHLHLQEAVLT	54	M	AC023167.7	60743 ... 61663
676	OR2CnP	2			FHGACAGTVGLMASFVL	59	M	M84005.1	1 ... 936
677	OR4TnP	9	14	0.21	MLSELLSHSQFVKSLSI	47	M	AC019272.4	62255 ... 61317
678	OR5BnP	1			FVITSGCNIHNIVVNDF	51	M	AF121977.1	262 ... 1197
679	OR4Kn	0	14	0.21	IHGGMTLHFQFINSISS	53	M	AB030896.1	1 ... 906
680	OR11Ln	0	1	254.43	LVGACVTTLHMILSVLI	50	M	AF121972.1	171 ... 1109
681	OR7E68P	5	10	17.21	MAGGELLDLHIMPAFGL	56	M	AF102536.1	22 ... 669
682	OR7EnP	2	10	17.21	MVACDVLDLHIIDSFGL	54	M	AF073989.1	547 ... 1515
683	OR7E31P	6	9	70.71	TAGGELLDLHIMPAFGL	55	M	AF073989.1	547 ... 1515
684	OR7EnP	3	9	70.71	MVACDVLDLHIMDSFGL	58	M	AF073989.1	547 ... 1515

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
685	OR5AKnP	3	11	52.82	LAATCGMNVHFLFVNLF	79	R	U50948.1	34 ... 978
686	OR5AKn	0	11	52.83	FAATCGMNVQFLFVNLF	79	R	U50948.1	34 ... 978
687	OR5AKn	0	11	52.83	FAATCGINVHFDFVDLF	79	R	U50948.1	34 ... 978
688	OR5BQnP	9	11	52.82	TTTTTLLLLLMLTFFFF	42	R	U50948.1	34 ... 978
689	OR1Nn	0	9	105.94	LLGGNVLPMHLIMGFLV	56	R	AF091566.1	1 ... 663
690	OR1J4	0	9	105.94	MITDNVLNSHLIVGVIL	69	M	AF102524.1	52 ... 669
691	OR1Nn	0	9	105.94	MLGDSLLVTHLVLGVLV	85	R	AB038167.1	1 ... 933
692	OR2AnP	4	3	94.41	TLAVCTIMVHHLGSIVG	65	M	AF102521.1	22 ... 669
693	OR2ANnP	17	9	93.78	......VVVLEFMVNLLI	23	M	AC074177.4	128803 ... 129726
694	OR5K1	0	3	104.47	FCETCGAHIHLLFSVQF	51	R	AF091575.1	52 ... 663
695	OR2K2	0	9	93.78	MLGSCVTTLEFMVSLLI	60	M	AJ251154.1	35662 ... 36615
696	OR8Hn	0	11	51.76	MAGTCGIDVNSIIVTLV	51	M	AC069559.8	36251 ... 35322
697	ORnP	15	11	51.76	LIFKNLFSPPLXXHYIL	28	M	X89682.1	2 ... 472
698	OR4AnP	14	11	50.28	FGRRVVGHIQLYGHNYV	38	M	AB030895.1	1 ... 924
699	OR4An	0	11	50.28	LHGGVVGQFQIVNGSCI	59	M	AB030895.1	1 ... 924
700	OR6Sn	0	14	0.58	FFGAFAGPGPADLAVIS	50	R	M64378.1	1 ... 933
701	OR4RnP	16	11	50.28	NLGAIMEHVXSVNGNYL	52	M	AF102522.1	40 ... 660
702	OR13Cn	0	9	86.77	MLGTCGINVQFLTTFLT	65	M	AJ133425.1	61 ... 1014
703	OR13DnP	4	9	86.77	MYGSCVLNTELIGNFLS	64	M	AC023789.5	371264 ... 372220
704	OR7EnP	3	11	2.13	MIACGVLDLHIINSFGL	54	R	AF091580.1	7 ... 663
705	OR10PnP	1	12	59.88	MIGICTTTTHLVATFII	49	M	AF247657.1	1 ... 945
706	OR8In	0	11	51.76	MVVCCMISISVSLATLS	50	M	AC069559.8	137090 ... 138039
707	OR8G1	0			..IIIGICVHCIVGNIV	75	R	AF091576.1	52 ... 663

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
708	ORnP	7	12	59.88	CFPGEAFFTLL......	34	M	AL359352.1	145887 ... 145042
709	OR5F1	0	11	51.76	MIATCGANVNHSLANIG	50	M	Y15525.1	1 ... 705
710	OR5FnP	1	11	51.76	MIATCGANVNYFFANKG	52	M	Y15525.1	1 ... 705
711	OR6BnP	6	2	251.7	LSVCCFSIIKFDLAILF	70	M	L14567.1	17 ... 667
712	OR2D1	0			LLGCCASVVDFITGILI	64	M	AF073987.1	2 ... 649
713	OR5ASn	0	11	51.76	MAADCLSTVHLLLCIQS	52	M	AC068904.15	165039 ... 165965
714	OR5SnP	8	2	251.7	FSSTTGRSVQLKLCMMN	64	R	AF091579.1	7 ... 663
715	OR5AQnP	0	11	51.76	SAVTDAGNTHGPFSIAF	51	R	X80671.1	203 ... 1129
716	OR6BnP	3	2	251.7	LSVCCFSIIKFDLAILF	67	M	L14567.1	17 ... 667
717	OR5JnP	2	11	51.76	YVLTGGGNTHGLFSIAL	52	R	X80671.1	203 ... 1129
718	OR9AnP	4	7	146.91	QLGTLVFFWPALMAIIG	44	M	NM_010991.1	1 ... 939
719	OR5BEnP	2	11	51.76	YSLTCVLNTHSFLSTST	45	R	AF091564.1	7 ... 663
720	OR9An	0	7	146.91	LLGTFVFFWPVLMAVLG	47	M	NM_010991.1	1 ... 939
721	OR8Hn	0	11	51.76	MVGTCGIDVNSIIATLV	51	M	AC069559.8	36251 ... 35322
722	OR5BNnP	14	11	51.76	LLMTCAYMSHS.....P	54	M	AF102528.1	52 ... 669
723	OR8Jn	0	11	51.76	LLIVVLYTVVCVSANLF	80	M	X89682.1	2 ... 472
724	OR9NnP	9	7	146.91	LFGTFIIIIIL.AAAAA	36	M	NM_010991.1	1 ... 939
725	OR7EnP	4	7		MVACGMLDLHITHSFAL	51	R	AF091580.1	7 ... 663
726	OR7E9P	3	7		MVACDVLDLHVIDSFGL	51	M	AF073989.1	547 ... 1515
727	OR8KnP	8	11	51.76	MMITLICQIIDILTNLP	36	M	AC069563.9	28460 ... 29383
728	OR2AnP	1	7	148.97	ILAHC............	44	M	AF102521.1	22 ... 669
729	OR8Kn	0	11	51.76	LLIIFIYQMFKSFSNLS	56	M	AF102528.1	52 ... 669
730	OR7E39P	4			MVGGELFHLHIMPAFGL	55	R	AF091580.1	7 ... 663

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
731	OR7E27P	3			MAGGELLDLHIMPAFGL	57	M	AF102536.1	22 ... 669
732	OR2Hn	0	6		FLGTCVMEVQSLASILV	81	M	AL078630.1	41097 ... 40165
733	OR13CnP	2	9	40.16	MLGACGATVQLMANFLV	87	M	AJ133428.1	61 ... 1017
734	OR13Cn	0	9	40.16	MFGACGAAVQLMTNFLV	89	M	AJ133424.1	61 ... 1017
735	OR2S1P	4	9	40.16	MFGACGANVQLMTNFLL	89	M	AJ251154.1	2703 ... 1747
736	OR2AMnP	1	9	40.16	.....RRRRRV.MMMMM	63	M	AJ251154.1	2703 ... 1747
737	OR1N1	0	1		MLGDSLLVTHLVLGVLV	85	R	AB038167.1	1 ... 933
738	OR2S2	0	9	40.13	MFAGCSIAVHLMTNFLV	83	M	AJ251154.1	2703 ... 1747
739	OR7E26P	4	1		MAGGELLDLHIMPAFGL	56	M	AF102536.1	22 ... 669
740	OR1F11	0			LAGNNGVNLHLIEGVMT	99	R	M64377.1	1 ... 939
741	OR5ACnP	3	3	103.97	FGATCIIHIHLIFSIQF	66	R	AF091575.1	52 ... 663
742	OR5B10P	2	13		MVATNGCNLRDLMSNVL	46	M	AF102528.1	52 ... 669
743	OR2AnP	1	12	85.7	TLAVCAFLVHLIACILG	76	M	AF102521.1	22 ... 669
744	OR1E5	0	13		MLGDSLLHLHLIMGILI	83	R	Y07557.1	1 ... 942
745	OR4Fn	0	6	185.71	IHGGMVLHFQFVNSICG	51	M	AB030896.1	1 ... 906
746	OR5CnP	0	9	40.53	MAADC............	47	M	Y15525.1	1 ... 705
747	OR2WnP	0	6	31.62	LLGGCVSNIMQALAIIA	64	M	AF102516.1	52 ... 669
748	OR2L2	0			..IIIGINAHYVSSFLL	48	M	AF102537.1	16 ... 669
749	OR4H8P	2	14		MHGCILGHVQLVNSISG	56	M	AF259072.1	104176 ... 105099
750	OR5D10P	5			LCVVTTWCTLFTSANES	44	R	AF010293.1	211 ... 1143
751	OR7A12P	1	14		MVIVSAMNIEMMSALGG	68	M	AF283558.1	1 ... 927
752	OR2L1	0			..IIIGINAHYVSTFLF	48	M	AF102527.1	22 ... 669
753	OR2F3P	0	14		LLGGFTSSVQIISSLLT	55	M	AF073974.1	41 ... 649

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
754	OR4H10P	2	15		MHGCILGHVQLVNSISG	57	M	AF259072.1	104176 ... 105099
755	OR5H1	0			..IIILGHIHFVFSIQF	56	R	AF091575.1	52 ... 663
756	OR2K1	0			..IIIITTLVCMVSLLI	58	M	AJ133428.1	61 ... 1017
757	OR7E11P	7	11		MAGGEFLDLHILPAFGL	52	M	AF073989.1	547 ... 1515
758	OR7A3P	1	11		MVIVSAMNIEMMSALGG	68	M	AF283558.1	1 ... 927
759	OR6A1	0	11		LLGCCGGIVKLDLAILG	91	R	M64386.1	130 ... 975
760	OR5I1	0	11		FCADSLGSVHFLYGVEI	52	M	Y15525.1	1 ... 705
761	OR2H3	0	6		ILGTCVIGVQSVASILV	86	M	AL078630.1	41097 ... 40165
762	OR10J1	0			MVGICGIVTQSTISVLV	73	M	X92969.1	8035 ... 8961
763	OR7E3P	3	11		MFACGVLDLHIIDSFGL	54	M	AF102536.1	22 ... 669
764	OR1D6P	1	11		LVVANLFYIHLLTGIFI	48	R	Y07557.1	1 ... 942
765	OR5D10P	2	18		LCVVTTWCTLFTSASES	45	R	U50948.1	34 ... 978
766	OR5D5P	2	18		LCVVTTWCTLFTSANES	46	M	AC073947.3	29192 ... 30115
767	OR52A1	0	11		MHQGSMAVCLIGVAVAF	72	M	NM_013620.1	1 ... 945
768	OR2AEn	0	7	98.36	HLGGCMGNIHIVSSLLL	48	M	AC073769.1	143294 ... 142353
769	OR6LnP	7	10	149.44	LLSSCSSAVSLRAAILA	40	M	NM_010983.1	178 ... 975
770	OR6LnP	7	10	149.44	LLSSCSSAVSLRAAILA	41	M	NM_010983.1	178 ... 975
771	OR7MnP	7	10	149.44	.....NVYVSL......	29	M	AC073947.3	43325 ... 42733
772	OR13Cn	0	9	86.77	MFGACGTDVQFMSNVLI	69	M	AJ133428.1	61 ... 1017
773	OR13Cn	0	9	86.85	MLGTCGANVQFMATFTM	71	M	AJ133425.1	61 ... 1014
774	OR2InP	6			LLGSC...........	79	M	AL078630.1	151152 ... 150391

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
775	OR4An	0	11	50.28	LHGGVVGHFQVVNSICV	58	M	AB030895.1	1 ... 924
776	OR2InP	3			......RRRRRMARILL	77	M	AL078630.1	151152 ... 150391
777	OR4AnP	4	11	50.28	LHGGVVGSFQVVNGICV	53	M	AB030896.1	1 ... 906
778	OR4AnP	7	11	50.28	PHGGAVAHFQVVNGICV	57	M	AB030896.1	1 ... 906
779	OR8C1P	2	11		LCVHCGMGVHCMIVVVV	72	M	AC068905.12	76922 ... 75948
780	OR4AnP	1	11	50.28	LHGDVVGHFQVVNGICV	56	M	AB030896.1	1 ... 906
781	OR7E15P	5	11		MAGGELQDVHIMPAFGL	54	M	AF073989.1	547 ... 1515
782	OR10A1	0	11		MFGVCAPVVQWAGTVVI	76	M	AF247657.1	1 ... 945
783	OR2An	0			TSAVCTCLVHLI.....	70	M	AF102521.1	22 ... 669
784	OR7EnP	6			MAGGELFHLHIMPAFGL	57	M	AF073989.1	547 ... 1515
785	OR7En	0			MAGGDFLDLHIVPAFVL	54	R	AF091580.1	7 ... 663
786	OR51A1P	5	11		MHTLSARLPLLAVITFL	43	R	AF079864.1	632 ... 1576
787	OR7E47P	4			KAGTNLLDLYIMPTFGL	56	M	AF073989.1	547 ... 1515
788	OR5B5P	2	3		MAATNICNIHELVANIS	48	M	AF146372.1	509 ... 1456
789	OR1F10	0	3		MFVDNGVNLHLIEGVMT	72	R	M64377.1	1 ... 939
790	OR8G2	0			..IIIGLGIHFVLSNIT	75	M	AF102518.1	52 ... 669
791	OR1Sn	0	11	54.08	MIVVNILITHLLVGVIF	55	M	AC073769.1	133488 ... 132556
792	OR4AnP	3	11	50.73	LHGGAVGHFQVVSGLCV	56	M	AB030896.1	1 ... 906
793	OR4AnP	7	11	50.76	LHGGILGHFQVVNGMCV	58	M	AB030896.1	1 ... 906
794	OR4AnP	5	11	50.66	LHGGVLGHFQVVNGMRV	56	M	AB030896.1	1 ... 906
795	OR4AnP	7	11	50.73	PHGGVVGRFQVVKVICV	54	M	AB030896.1	1 ... 906
796	OR4AnP	1	11	50.81	LHGGIVGHFQVVSGMCV	60	M	AB030896.1	1 ... 906
797	OR4AnP	10	11	50.81	LHGGVVGNFQVVNGICV	55	M	AF102522.1	40 ... 660
798	OR4An	0	11	50.73	LHAGVAGHVQFMNGICV	62	M	AB030895.1	1 ... 924
799	OR4An	0	11	50.73	LHGGVVGHVQFVNGICV	57	M	AB030896.1	1 ... 906
800	OR7E42P	4			MAGGELQDVHIMPAFGL	54	M	AF073989.1	547 ... 1515

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
801	OR2M3P	2			ITLGCFLDIDALCCMIF	55	M	AF102537.1	16 ... 669
802	OR4H11 P	2	4		MHGCILGHVQLVNSISG	57	M	AF259072.1	104176 ... 105099
803	OR7E57 P	5			MAXGEFLDLHILPAFGL	51	M	AF102536.1	22 ... 669
804	OR2B1P	0	5		LLGAYATNWLLLVSFHI	78	R	L34074.1	73 ... 1011
805	OR7E34 P	2			MAGGDSLDLHIMPAFGL	56	M	AF073989.1	547 ... 1515
806	OR7E56 P	4			MAGDELFFLHILPAFGL	52	M	AF073989.1	547 ... 1515
807	OR3AnP	1	5		LHAGCACNTHALAAMAA	49	M	AF073967.1	2 ... 649
808	OR4H5P	2	5		MHGCILGHVQLVNSISG	56	M	AF259072.1	104176 ... 105099
809	OR1En	0	5		MLGDSLLHLHLIMGILI	82	R	Y07557.1	1 ... 942
810	OR51Cn P	2	11	3	MKTVSYYYIXQ......	48	M	AF121975.1	50 ... 1012
811	OR2WnP	2	6	30.51	LLGGCVSNIMQALAIIA	64	M	AF102516.1	52 ... 669
812	OR51B1 P	5	11		AHSVSGRSPVRPLITIL	68	M	AF071080.2	15931 ... 16851
813	OR7E81 P	3			MAGGEFFSLHIMPAFGL	54	M	AF102536.1	22 ... 669
814	OR7E44 P	1			MAGGELFDLHIMLAFGL	53	M	AF073989.1	547 ... 1515
815	OR5B7P	2	6		MAATNICNIHELVANIS	47	M	NM_013728. 1	1 ... 948
816	OR7E36 P	4			MAGGELFFLHIMPAFGL	58	M	AF073989.1	547 ... 1515
817	OR2A5	0	7		TMAHCTCLVHLIASILG	74	M	AF102521.1	22 ... 669
818	OR5B1P	2	8		MAATNICNIHELVANIS	47	M	AF146372.1	509 ... 1456
819	OR8B8	0	11	137.68	LLVVSGMGAHCVVVDIV	72	M	AC069559.8	120212 ... 119283
820	OR8B4P	0	11	137.71	LCVNCGVGAHSFVVITL	87	M	AC068910.2 1	133103 ... 132162

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
821	ORnP	15	11	137.77	LCVENRRTATHCKSHII	35	M	AC069563.9	60295 ... 59327
822	OR8B3	0	11	137.77	LLVICAMGAHCVVVNIV	85	M	AC069563.9	129775 ... 130725
823	OR2Bn	0	6	30.51	LLGSCASNLQWLISFLI	89	R	L34074.1	73 ... 1011
824	OR8B6P	6	11	137.77	LAFFCGLSAHCVAAAVI	73	M	AC069559.8	96224 ... 95292
825	OR8B5P	6	11	137.77	LFFFXGLGAHCVVANTV	73	M	AC069559.8	96224 ... 95292
826	OR4E2	0	14	1.7	LHACIAGHGQLINSISS	90	M	AF259072.1	104176 ... 105099
827	OR8B7P	4	11	137.77	FCVICGWGAHCVAAIFV	71	M	AC069559.8	96224 ... 95292
828	OR11JnP	3	15	1.82	FSCAGFGSMPLCVSIII	56	M	AF121972.1	171 ... 1109
829	OR4E1P	3	14	1.7	MHACIAGHALLINSISV	92	M	AB030893.1	37 ... 930
830	OR10DnP	7	11	137.96	.....HHHILLGNVLSI	85	M	AC074177.4	12106 ... 13038
831	ORnP	10	14	1.7	VFRGGFHKFFF......	23	M	AF102536.1	22 ... 669
832	OR8D2	0	11	137.77	LLVIGVLWVHRLIGNTA	70	M	AC073947.3	29192 ... 30115
833	OR11InP	1	1	126.31	FGAACGCLITLATSVTI	51	M	AL359381.1	175785 ... 176720
834	OR11JnP	1	15	1.82	FSCACFGWTPLCISIIL	56	M	AF121972.1	171 ... 1109
835	OR10AnP	3	11	5.64	MFGVCTPVVQWAGTVVI	74	M	AF247657.1	1 ... 945
836	OR8C3P	5	11	137.77	LCVHCGMGVHCMIVVVV	73	M	AC068905.12	76922 ... 75948
837	OR2DnP	6	11	5.64	LLGCCGSVVDFITGILI	62	M	AF073987.1	2 ... 649
838	OR4PnP	0	11	51.03	LHGGIVGHSQL......	59	M	AB030895.1	1 ... 924
839	OR7E21P	5			MAGGEFIDLHIMPAFGL	50	M	AF073989.1	547 ... 1515
840	OR2M1	0			IVLGCFLDIYAICSMLF	55	M	AF102537.1	16 ... 669
841	OR7AnP	4	19		NLAGVVMNLQM......	63	M	AF073970.1	41 ... 649

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
842	OR5D11 P	1	8		LCVVTTWCTLFTSANES	44	R	AF010293.1	211 ... 1143
843	OR7E50 P	7	8		IVVCDMLDLHVFLDIFL	57	M	AF102536.1	22 ... 669
844	OR7E45 P	3			MAGGELFDLHIMPAFGL	54	M	AF073989.1	547 ... 1515
845	OR7E77 P	6			MAGGEFLDLHIMPAFGL	51	M	AF073989.1	547 ... 1515
846	OR8B2	0	11	137.77	LLVICAMGAHCVVVNIV	84	M	AC069563.9	129775 ... 130725
847	OR8D1	0	11	137.77	LVVVGALSTHALIANTV	87	M	AC073947.3	29192 ... 30115
848	OR8B1P	4	11	137.77	LLLVCGMGAHCVVVNIV	84	M	AC069559.8	96224 ... 95292
849	OR7A1P	2	19		MIVVSVVYLQMMTSLGG	72	R	M64376.1	1 ... 999
850	OR7E8P	4	8	13.72	MVACGVLDLHIIDSFGL	53	M	AF102536.1	22 ... 669
851	OR4DnP	7	11	55.86	MHGGVAGHVQLMNNISL	58	M	AC019272.4	183633 ... 182701
852	OR7E80 P	7	8	13.72	MAGGELQDVHIMPAFGL	54	M	AF073989.1	547 ... 1515
853	OR4DnP	5	11	55.86	MHGGAAGHVQLMNNLTL	62	M	AC019272.4	183633 ... 182701
854	OR7E10 P	8	8	13.72	IVACDLLDLHIIDSFGL	55	M	AF073989.1	547 ... 1515
855	OR10B1 P	3	19	17.91	MLGCCLSVIEMILSVVM	85	M	AC012302.5	54283 ... 55224
856	OR2InP	3			.......LLLLMARILL	75	M	AL078630.1	151152 ... 150391
857	OR4Dn	0	11	55.86	MHGGVGGHAQLMNNVSF	65	M	AC019272.4	183633 ... 182701
858	OR5ACn	0			.VVVVIIHVHLIFGIQP	65	R	AF091575.1	52 ... 663
859	OR2I1	0	6	33.63	LLGSCASNAQLMARILL	79	M	AL078630.1	151152 ... 150391
860	OR10H1	0	19	19.86	MFGFSCGMVVAGLVTAL	88	M	AC023604.2	245345 ... 246298

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
861	OR7E59 P	5			CPEARVFLLHIMPAFGL	53	M	AF102536.1	22 ... 669
862	OR7E28 P	4			MAGGELLDLHIMPAFGL	54	M	AF073989.1	547 ... 1515
863	OR5B3	0			MVATNGCNIHDLVVNII	51	R	U50948.1	34 ... 978
864	OR2A6	0			TLAHCAFLVPLIACILG	75	M	AF102521.1	22 ... 669
865	OR6Cn	0			.VVVVCAIPPLVMAALI	47	M	NM_010991.1	1 ... 939
866	OR7E54 P	5			MAGGEFLDLHIMPAFGL	52	M	AF073989.1	547 ... 1515
867	OR7E48 P	3			MAGGEFLDLHIMPAFGL	57	R	AF091580.1	7 ... 663
868	OR67An P	3	11	76.42	MHSCAGTLPAQGIAVSL	83	R	AF091561.1	52 ... 663
869	OR4DnP	1	11	55.86	MHGGVAGHVQLMNNLTL	63	M	AC019272.4	183633 ... 182701
870	OR4CnP	1	11	50.91	VHGCILGHAQLLNSICS	57	M	AB030896.1	1 ... 906
871	OR4DnP	2	11	55.86	IHGGIAGHVQLMNNVTL	65	M	AC019272.4	183633 ... 182701
872	OR10H2	0	19	19.94	MFGFSCGMVVAGLVMAL	85	M	AC023604.2	245345 ... 246298
873	OR10H3	0	19	19.94	MFGFSWGMMVMGLVTAI	75	M	AC023604.2	214343 ... 213396
874	OR55Cn P	2	11	2.65	VYLLYLQPGGG......	45	M	AF121980.1	160 ... 1053
875	OR55Bn P	3	11	2.65	.VVVVLQVPLLGMCTVS	53	M	AF121980.1	160 ... 1053
876	OR52Vn P	4	11	4.19	LHNHIMVYXFLGTTSPL	48	M	NM_013619.1	118 ... 969
877	OR2B3	0	6	33.64	LLGACFINLQLLFSILI	75	R	L34074.1	73 ... 1011
878	OR52Tn P	6	11	4.22	FGHFLIFLDFLDILTIS	45	M	AF121975.1	50 ... 1012
879	OR2J1P	5	6	33.64	LLGTCASTLHFLMSFVI	57	R	L34074.1	73 ... 1011
880	OR52Hn P	3	11	4.19	LHFVSGRVPCLGVPTVT	60	M	AF121975.1	50 ... 1012

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
881	OR2J3	0	6	33.64	LLGTCASNLHFLTSFVI	58	R	L34074.1	73 ... 1011
882	OR52An	0			FHSVS......VVRLFS	75	R	AF079864.1	632 ... 1576
883	OR4Qn	0			.VVVVAGHMQLVNSLSV	56	M	AB030893.1	37 ... 930
884	OR52Bn P	2	11	4.22	LHFVSVRTSILGVPSVL	60	M	AF121975.1	50 ... 1012
885	OR2N1P	9	6	33.64	LHGGCPIYSEALVCMLV	81	M	AJ132195.1	79 ... 906
886	OR51En P	1			FHSASVRFPLLGAIAMV	90	R	AF079864.1	632 ... 1576
887	OR2J2	0	6	33.64	LLGICAIILHFLMSFVI	57	R	L34074.1	73 ... 1011
888	OR2In	0			.....RRRRRRMARILR	77	M	AL078630.1	151152 ... 150391
889	OR2J4P	5	6	33.64	LLGTCASNLHFLTSFVL	56	R	L34074.1	73 ... 1011
890	OR7E40 P	4			MAGGDILDLYILPDFGL	55	M	AF073989.1	547 ... 1515
891	OR2H4P	3	6	33.64	LLGAYLTQIQAMASLLM	63	M	AL078630.1	41097 ... 40165
892	OR7E52 P	5			IVVCDVLDLHVCDIFGL	61	M	AF073989.1	547 ... 1515
893	OR2InP	9			LLGSC............	80	M	AL078630.1	151152 ... 150391
894	OR6C1	0			LIGVFTVIPALGCATLF	52	M	NM_010991.1	1 ... 939
895	OR7E30 P	3			MAGGEFLDLHIMPAFGL	56	M	AF073989.1	547 ... 1515
896	OR5BAn P	0	11	53.69	LVVTSVFNIQNLFSVTL	51	R	AF091579.1	7 ... 663
897	OR7H1P	3	19	11.38	MMGGTVLYIQLLVALDV	74	M	AF073989.1	547 ... 1515
898	OR5B2	0	11	54.45	MVATNGCNFHGLTSNIF	47	R	U50948.1	34 ... 978
899	OR5AZn P	1	11	53.69	MIGTCTVNLLCILCLIF	48	R	AF091579.1	7 ... 663
900	OR5Bn	0	11	54.45	MVATNGCNIHDLVVNII	51	R	U50948.1	34 ... 978

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
901	OR52Bn	0	11	4.22	KILFSARIPSLGAASTL	64	M	NM_013619.1	118 ... 969
902	OR5BnP	2	11	54.45	MAATNICNIHELVANIS	49	R	U50948.1	34 ... 978
903	OR52Dn	0	11	4.19	MHYASVRIPFLGVAAML	66	M	AF121976.2	474 ... 1307
904	OR7A11	1	19	17.72	MVEASAIDLHMMAVLGV	67	M	AF283558.1	1 ... 927
905	OR5BnP	9	11	54.45	MAATSALTVDDLLQFFL	41	M	NM_013728.1	1 ... 948
906	OR51AnP	5	11	4.19	THSWFSRMPLLGIVAFV	50	R	AF079864.1	632 ... 1576
907	OR7A15P	4	19	17.72	MIVGSVTHLHMMAALGG	74	R	M64376.1	1 ... 999
908	OR7C2	0	19	17.72	IIGCNGIGLETMVTLGF	98	R	AF091580.1	7 ... 663
909	OR7E23P	7	21	20.89	MAGGELFHLQIMPAFGL	57	M	AF073989.1	547 ... 1515
910	OR2E1	8	6	32.05	AHACCTINLQI.RRRRR	43	M	AL078630.1	106872 ... 105934
911	OR1I1	0	19	17.87	MHGTSAIQIHLIFGVGS	57	R	AF091566.1	1 ... 663
912	OR1RnP	3	17	3.12	MVGISAVHLHLIEGVVA	45	R	M64377.1	1 ... 939
913	OR4F3	0	8	0.07	IHGGMVLHFQFVNSICG	51	M	AB030896.1	1 ... 906
914	OR2AEn	0	7	98.7	HLGGCMGNIHIVSSLLL	49	M	AC073769.1	143294 ... 142353
915	OR2InP	7			......TTTTTMARILL	72	M	AL078630.1	151152 ... 150391
916	OR52AnP	2			IHSASVRFPLLGXPPPP	94	R	AF079864.1	632 ... 1576
917	OR7C1	0	19		ITGCNGIGLETIATLGI	81	R	AF091580.1	7 ... 663
918	OR2A3P	2	7	149.11	MLAACTCLINLVGGVLG	63	M	AF102521.1	22 ... 669
919	OR7A5	0	19		MIAGNAMYLQMITVLGG	74	M	AF283558.1	1 ... 927
920	OR2InP	3			...........MARILL	67	M	AL078630.1	151152 ... 150391
921	OR7A10	0	19		MLVGNAMNLQMMAVLGG	76	R	M64376.1	1 ... 999
922	OR2An	0			....	81	M	AF102521.1	22 ... 669
923	OR2M2	0			IISGCFLDIDAICCMLF	57	M	AF102537.1	16 ... 669

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
924	OR7A8P	2	19		MLAVSSLNLQMIATLGG	71	M	AF283558.1	1 ... 927
925	OR2An	0			TSAVCTTLIHL......	78	M	L14566.1	62 ... 667
926	OR7E20P	4			MAGGELLFLHIMPAFGL	56	M	AF073989.1	547 ... 1515
927	OR2AnP	3			TLAHCTCLVHL......	65	M	AF102521.1	22 ... 669
928	OR5BHnP	7			MVASCGGKTVS......	34	M	Y15525.1	1 ... 705
929	OR1En	0			LMGDSLLHLHLIMGISI	92	M	AC068902.11	196434 ... 195499
930	OR1EnP	1			MLGDSLLHLHLIIGVVL	98	M	AF073976.1	32 ... 649
931	OR5Bn	0	11	54.45	FVITSGCNIHNIVVNDF	51	R	U50948.1	34 ... 978
932	OR8RnP	12	11	73.74	LFLSYGGGAHH......	52	M	AC069561.10	7848 ... 8783
933	OR5ANn	0	11	55.69	YSGLSGTAFQATLTFGA	55	R	AF091564.1	7 ... 663
934	OR5ANnP	1	11	55.69	YSGLCGTGIQATLTFGT	59	M	Y15525.1	1 ... 705
935	OR5BRnP	8	11	55.69	MSNVCGTVIQATLTFGT	33	M	Y15525.1	1 ... 705
936	OR2A1	0	7	149.18	TLGHCTCLAHLIACFLG	77	M	AF102521.1	22 ... 669
937	OR10An	0	11	6.81	MLGGCFLLVQWAGTIIV	54	M	AF247657.1	1 ... 945
938	OR2A9	3	7	149.18	TLAHCTCLVHLIACILG	78	M	AF102521.1	22 ... 669
939	OR2A7	0	7	149.18	TSAVCTTLIHLVGAGLG	81	M	L14566.1	62 ... 667
940	OR10A3	0	11	6.81	MLGGCFSVVQWAGTIVV	58	M	AF247657.1	1 ... 945
941	OR10Cn	0	6	33.36	MLGACSCVGHFIATLIC	59	M	AL365336.1	122764 ... 121784
942	OR7A2P	0	19		MVIVSVMNLQVMAALDG	73	M	AF283558.1	1 ... 927
943	OR10WnP	2	11	54.3	MIGSCASLQLFVAAAIV	47	M	AC012302.5	54283 ... 55224
944	OR7A17	0	19		MVGGSAINSQMMAALAG	76	M	AF283558.1	1 ... 927
945	OR5Bn	0	11	54.3	MAATNGINIQDLISNVF	47	M	AF102528.1	52 ... 669
946	OR5BnP	5	11	54.3	MVATNGCNLRDLMSNVL	47	M	AF102528.1	52 ... 669

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
947	OR1Q1	0	9	106.13	TIAVNMLHLHLIEGVIG	54	M	AF073967.1	2 ... 649
948	OR2Hn	0	6	33.33	LLGTCVMQVQSLSSFVV	88	M	AL078630.1	48786 ... 47851
949	OR7EnP	5	3	90.04	MVACDVLDLHIIDSFGL	54	M	AF073989.1	547 ... 1515
950	OR7A14	0	19	17.72	MVIVSAMNI........	71	M	AC073772.1	227187 ... 226252
951	OR1B1	0	9	106.13	FYGVTLVHLRLIEGLMG	49	M	AC068902.1 1	83719 ... 84647
952	OR12D2	0	6	33.23	LHGSSTIHLHMLVTIAG	81	M	AL359381.1	105330 ... 104407
953	OR7EnP	4	3	11.92	MVACDVLDLHIIDSFGL	55	M	AF073989.1	547 ... 1515
954	OR8BnP	5	15	74.31	LXVVEGMGAHCVVVNIV	82	M	AC069559.8	96224 ... 95292
955	OR1L1	0	9	106.13	MLGNSLIHLHLVEGVIT	57	M	AC023167.7	60743 ... 61663
956	OR11An	0	6	33.36	FGATCTSVLVLTLSCLI	76	M	AL359381.1	175785 ... 176720
957	OR7AnP	4	12	44.29	....HLLDCYIRTTLSG	55	M	AF102534.1	52 ... 669
958	OR1C1	0	1	254.35	LVVNSGVHLHLIVGLAT	56	M	AC073769.1	133488 ... 132556
959	OR1D2	0	17	2.99	LVVANLLYIHLLTGIFI	50	M	AF073967.1	2 ... 649
960	OR1L3	0	9	106.13	MLGNSFFHLHLAEGSVA	53	M	AC023167.7	14677 ... 15636
961	OR12Dn P	1	6	33.36	LHGSATIHLHMSTGIAG	76	M	AL359381.1	105330 ... 104407
962	OR4G1P	4	16	83.04	KHGGMAIHSQFVNSISG	47	M	AB030896.1	1 ... 906
963	OR2B4P	1	6	33.53	LLGSCGSNVQLLLGLLM	90	M	AL359352.1	95024 ... 95965
964	OR11H1	0	22		FFGTCLCWIPLCLSVIG	61	M	AC027184.3	54955 ... 54017
965	OR4Fn	0	16	83.04	IHGGMVIHSQFVNSLTC	50	M	AC019272.4	62255 ... 61317
966	OR56An P	5	11	4.73	MNLPSFQLPVLQAGFLS	38	M	AF121975.1	50 ... 1012

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
967	OR8NnP	7	4	164.13	REIIRVDAFLKKTANMI	34	M	AF102528.1	52 ... 669
968	OR7EnP	5			MVACDVLDLHIFFDFGL	54	R	AF091580.1	7 ... 663
969	OR4Pn	0	11	50.95	LHGGIVGHSQLVNSIAV	56	M	AB030895.1	1 ... 924
970	OR6Cn	0			LIGVFCSTPPLGFATLF	51	M	NM_010991.1	1 ... 939
971	OR5BCnP	2	11	54.3	.....GCQIHFLLANIF	41	M	AC069561.10	51687 ... 50743
972	OR10QnP	4	11	54.3	MLGGCGLLQLLLVSVLV	48	M	AC012302.5	54283 ... 55224
973	OR5BnP	6	11	54.3	TDASNGGNIHELVTNIF	45	R	U50948.1	34 ... 978
974	OR10PnP	2	12	115.61	MIGICTTTTHLVATFII	46	M	AF247657.1	1 ... 945
975	OR1L4	0	9	106.22	MMGNSGIHFRLVETVIT	62	M	AF073967.1	2 ... 649
976	OR2APnP	3	12	115.61	YMGAFLLLLLL......	49	M	AF073987.1	2 ... 649
977	OR1L6	0	9	106.22	MMGNSGIHFRLVETVIT	63	M	AF073967.1	2 ... 649
978	OR6UnP	6	12	115.61	DIGAFTLFMPLDLAALG	52	M	NM_010991.1	1 ... 939
979	OR5C1	0	9	106.06	MAADCAGSVHLLICIQA	50	R	X80671.1	203 ... 1129
980	OR11InP	1	15	70.72	FGAACGCLITLATSVTI	51	M	AL359381.1	175785 ... 176720
981	OR4AnP	6	11	50.78	LYGGVVGHFQVVNGVCV	57	M	AB030896.1	1 ... 906
982	OR4GnP	14	2	114.45	ICRKMAVHSQFVNSISA	42	M	AB030892.1	1 ... 939
983	OR10Vn	0	11	56.15	MVGGCGLLPLLLISVLI	48	M	AL136158.14	29455 ... 30402
984	OR4G2P	2	2	114.45	KHGGMAIHSQFVNSISG	48	M	AB030896.1	1 ... 906
985	OR10VnP	3	11	56.15	MIGRCGLLQLLMVSFLV	45	M	X92969.1	8035 ... 8961
986	OR4F4	0	2	114.45	IHGGMVIHSQFVNSLTC	50	M	AC019272.4	62255 ... 61317
987	OR4G3P	14	19	63.51	ICRKMAVHSQFVNSISA	42	M	AB030892.1	1 ... 939
988	OR5AKnP	4	11	52.82	LGATCSMNINFLFVNLC	65	R	U50948.1	34 ... 978
989	OR10YnP	14	11	56.15	MIRGCGLLFLLLCGHHL	43	M	AF247657.1	1 ... 945
990	OR4GnP	2	19	63.51	KHGGMAIHSQFVNSISG	48	M	AB030896.1	1 ... 906

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
991	ORnP	9	5	111.92	IMCSRTTYVXQLHGFFT	23	M	AF073989.1	547 ... 1515
992	OR4Fn	0	19	63.51	IHGGMVIHSQFVNSLTC	50	M	AC019272.4	62255 ... 61317
993	OR8A1	0	11	137.56	LLVICVIGIELVSANIV	61	M	AC069559.8	96224 ... 95292
994	OR8Bn	0	11	137.56	LCVVSGMGAHSVVVDVM	66	M	AC069559.8	120212 ... 119283
995	OR6DnP	3	10	47.91	AYVSSLLLRTH......	55	R	AF034901.1	2110 ... 3078
996	OR7E14 P	7	11	16.31	MAGGELLDLHIMPAFGL	58	R	AF091580.1	7 ... 663
997	OR2M4	0			IVLGCALDIVALCCMLF	57	M	AF102537.1	16 ... 669
998	OR4WnP	3	X		LLLLL......LLFFII	36	M	AC069559.8	73704 ... 74636
999	OR4Fn	0	19	63.51	IHGGMVIHSQFVNSLTC	50	M	AC019272.4	62255 ... 61317
1000	OR7EnP	3			MAGGESLDLHIMPAFGL	57	M	AF073989.1	547 ... 1515
1001	OR4GnP	4	19	63.51	KHGGMAIHSQFVNSISG	47	M	AB030896.1	1 ... 906
1002	OR10Jn P	1			LLGVCGITIQSTISVLL	60	M	X92969.1	8035 ... 8961
1003	OR52En	0	11	4.58	MHTASIRMPLLGNILLL	71	M	AF121979.1	53 ... 1106
1004	OR4RnP	24	11		VHGAIMGHVXSFANNCL	54	M	AF102522.1	40 ... 660
1005	OR4Cn	0	11		AHGAIVGHIQFVNSICL	75	M	AF102522.1	40 ... 660
1006	OR4AnP	10	11		GLGGIVGHIQL......	44	M	AF102522.1	40 ... 660
1007	OR4AnP	4	11		LHGGVAGHFQVVNGGCI	55	M	AB030895.1	1 ... 924
1008	OR4AnP	8	11		LHGGVAGHSHSVNGICV	54	M	AF102522.1	40 ... 660
1009	OR9Gn	0	11	52.54	FAAYCVGNIIKMLLNVC	46	M	AC074177.4	106297 ... 105361
1010	OR10An	0	12	59.65	MFGSCGSVLQWASTFIF	64	M	AF247657.1	1 ... 945
1011	OR4Cn	0	11		VHRGVVGHIQFINSICL	73	M	AF102522.1	40 ... 660

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
1012	OR10VnP	8	11	56.15	.FFFFIIXNEXSVVVLV	37	M	AC073945.4	110931 ... 111893
1013	OR10UnP	3	12	59.65	MAGLCATVAQLMLSFIS	56	R	AF034898.1	1 ... 981
1014	OR7E2P	3	11	90.37	MVACDVLDLHICDIFGL	59	M	AF073989.1	547 ... 1515
1015	OR7E35P	6	4	11.87	MAGGEFLDLHIVPAFVL	53	M	AF102536.1	22 ... 669
1016	OR9KnP	0	12	59.71	LAIVGGCSLQVSLSIIP	49	R	AF091579.1	7 ... 663
1017	OR7E13P	5	11	90.37	MAGGEFLDLHIMLAFGL	54	R	AF091580.1	7 ... 663
1018	OR7EnP	4	8	6.5	MLACGVLDLHIIDSFGL	55	M	AF102536.1	22 ... 669
1019	OR9Kn	0	12	59.71	LAIVGGCSIQMSLSIIP	49	M	NM_013728.1	1 ... 948
1020	ORnP	13	11	137.56	PCVIYGIDVHSLXEPAY	34	M	AC069559.8	36251 ... 35322
1021	OR7EnP	8	11	72.11	MAGGNLFFSLLMPAFGL	54	M	AF073989.1	547 ... 1515
1022	OR7EnP	5	3	140.64	MAGGKFLDLHIMPAFGL	53	M	AF073989.1	547 ... 1515
1023	OR3A4P	0	17	3.12	LHAGCMFNTQALAAMGA	44	M	AC073769.1	133488 ... 132556
1024	OR8QnP	9	11	137.56	LSIIIVETEFVFTXIVT	33	M	AC069559.8	137090 ... 138039
1025	OR7EnP	2	11	72.11	ILACGVLDLHIMHNFGL	55	M	AF073989.1	547 ... 1515
1026	OR7EnP	3	3	140.64	MVACGVLDLHIIHSFGL	56	M	AF073989.1	547 ... 1515
1027	OR3A1	0	17	3.07	LHVGCACNTHALVGMAT	50	M	AF073967.1	2 ... 649
1028	OR5Gn	0	11	52.52	MGEACGMSTHFLLAIGL	69	M	AF146372.1	509 ... 1456
1029	OR5MnP	7	4	42.45	LIIIYVYNAQRIIIMLE	39	M	AF073987.1	2 ... 649
1030	OR7EnP	1	3	136.02	MVACDVLDLHIIDNFGL	54	M	AF073989.1	547 ... 1515
1031	OR5G1P	2	11	52.51	QGVACGINTHNVVAVGF	68	M	AF146372.1	509 ... 1456
1032	OR5PnP	3	11	6.93	LVGTCAGNSFCPSSVLS	70	M	AF121977.1	262 ... 1197

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
1033	OR10AEnP	8	1	157.36	IIIIIGIMVIVQIHCVV	40	M	X92969.1	8035 ... 8961
1034	OR3A2	0	17	3.07	LHAGCACNTHALVGMAT	50	M	AC073769.1	133488 ... 132556
1035	OR10Jn	0	1	157.4	MVATCGIMLHANVSVIV	88	M	X92969.1	8035 ... 8961
1036	OR1D3P	2	17	2.94	LVVANLFYIHLLTGIFI	50	R	Y07557.1	1 ... 942
1037	OR10Jn	0	1	157.36	TVAICGIMVQSNVRVIV	72	M	X92969.1	8035 ... 8961
1038	OR1D4	0	17	2.99	LVVTNLLYLLLLTGIFT	49	R	Y07557.1	1 ... 942
1039	OR5GnP	8	11	52.51	QGVVYVANTHAVVAVLV	55	M	NM_013728.1	1 ... 948
1040	OR4SnP	1	11	50.99	LHGCIGGHIQLVNSIAG	61	M	AB030895.1	1 ... 924
1041	OR5GnP	4	11	52.51	LGVVCGVSTHFLLVLGL	75	M	AF146372.1	509 ... 1456
1042	OR9HnP	2	1	254.35	FSGIAGWNAQMLLCIIS	59	R	AF091579.1	7 ... 663
1043	OR1A1	0	17	2.99	MIGNSGINPHLMGVIFV	86	M	AF073966.1	41 ... 643
1044	OR1A2	0	17	2.99	MIAKSGISPHLMLGVFL	80	M	AF073966.1	41 ... 643
1045	OR8AnP	6	11	137.68	FLVICVMVIELVFANLI	50	M	AC069561.10	51687 ... 50743
1046	OR1P1P	1	17	2.99	LLGDIALLTRLLLGVII	82	M	AF102538.1	139 ... 675
1047	OR7E12P	7	11	1.92	MAGGEFFSLHIMPAFGL	55	M	AF073989.1	547 ... 1515
1048	OR4A1P	4	11		LHGGVVGHFQVVNGICV	57	M	AB030896.1	1 ... 906
1049	OR10G3	0	14	1.7	LHGSCGAHLQLTDIVVS	91	M	AF259072.1	19582 ... 18644
1050	OR10G1P	3	14	1.7	LHGSCGAHIQLTDIVAS	93	M	AF259072.1	55611 ... 54658
1051	OR10G2	0	14	1.7	LHGSCGAHIQLTDVVAS	91	M	AF259072.1	55611 ... 54658
1052	OR5Tn	0	11	51.94	MVGTCAAHIHALFVIEV	52	M	AF121977.1	262 ... 1197
1053	OR7EnP	8	3	136.02	MVACGVLDLHIIGSFGL	53	R	AF091580.1	7 ... 663
1054	OR7EnP	5	3	136.02	MAGGKFLDLHIMPAFGL	54	M	AF073989.1	547 ... 1515
1055	OR4AnP	2	11	50.93	LHAGVVGHVQFMNGICV	61	M	AB030895.1	1 ... 924
1056	OR4C1	1	11	50.93	LHGGIIGHVQFVNSMCL	66	M	AB030896.1	1 ... 906

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
1057	OR1EnP	7	17	2.9	.....MMMYTLIMGILI	80	M	AF073961.1	32 ... 649
1058	OR7KnP	11	14	5.99	MIGCNFIELYMMIGIFG	49	R	AF091580.1	7 ... 663
1059	OR4CnP	3	11	50.93	LHDGIEGHIQFVNSMCA	61	M	AF102522.1	40 ... 660
1060	OR1RnP	11	17	2.9	MVGISAVHLHLIEGVVA	44	R	M64377.1	1 ... 939
1061	OR5AUn	0	14	1.22	MAATCGANIHCLFANLS	51	M	AC069559.8	85584 ... 84655
1062	OR4Cn	0	11	50.96	LHAGVVGHIQFVNSICI	69	M	AF102522.1	40 ... 660
1063	OR4Cn	0	11	50.96	VHGCIVGHVQLLNSICV	57	M	AB030895.1	1 ... 924
1064	OR13DnP	2	9	86.89	MLGSCWITLRLFTVIVL	58	M	AJ251154.1	2703 ... 1747
1065	OR5n				ASASLTSYVHNEEEVFV	44	M	AL359352.1	111313 ... 112242
1066	OR2Hn				LLGTCVMQVQSLSSLVV	83	M	AL078630.1	48786 ... 47851
1067	ORn				....	25	M	AC074177.4	88434 ... 88916
1068	ORn				.....EINLLLARGKAL	29	M	AF283814.1	1 ... 930
1069	ORn				NNNNNFXSLHLCCCILI	29	M	AC074177.4	128803 ... 129726
1070	ORn				TLLLLTFQHHL......	27	M	L14569.1	62 ... 667
1071	OR6Fn				..CCCWPIPTSAIAVIS	46	R	M64386.1	130 ... 975
1072	ORn				ILLLLL	33	R	U50947.1	418 ... 1350
1073	ORn				..CCCLIPFFFTSGYSW	24	R	M64392.1	1 ... 942
1074	OR10An				PLGECDPEEQMYVGLVM	51	M	AF247657.1	1 ... 945
1075	ORn				IPNASRRRRRR....PP	25	R	M64388.1	1 ... 942
1076	OR2Ln				FLAGAGINAHYVSTFLF	51	M	AF102527.1	22 ... 669
1077	OR10Jn				LTGICGIMVQSNVSVLL	57	M	X92969.1	8035 ... 8961
1078	OR1Kn				LLLLLMVNLYLIKGVVT	50	R	M64377.1	1 ... 939
1079	OR10Dn				LHGSCGLHILLSNVISG	69	M	AC074177.4	12106 ... 13038
1080	ORn				CCCIII	41	R	M64376.1	1 ... 999

SEQ ID #	Symbol	D	C	Mb coord	CDR	%	S	Acc	Range
1081	OR2Ln				SLACGGLNAHFVRTLSF	52	M	AF102537.1	16 ... 669
1082	ORn				HHHHHRLESSSLLLLLL	38	M	AC073945.4	152209 ... 153150
1083	ORn				...........LLLLLS	27	M	AL365336.1	41087 ... 41711
1084	OR2n				...........GGGGGG	57	M	AF102521.1	22 ... 669

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be apparent to those skilled in the art that various changes and modifications can be practiced without departing from the spirit of the invention. Therefore the foregoing descriptions and examples should not be construed as limiting the scope of the invention.

All patents, patent applications, and publications cited herein are hereby incorporated by reference in their entirety. In particular, the following documents are hereby incorporated by reference in their entirety: United States Provisional Patent Applications Serial Nos. 60/145,412, filed July 23, 1999; 60/155,126, filed September 22, 1999; 60/158,495, filed October 8, 1999; 60/158,615, filed October 8, 1999; 60/181,113, filed February 8, 2000; 60/181,115, filed February 8, 2000; 60/184,809, filed February 24, 2000; 60/188,332, filed March 9, 2000; and United States Patent Applications Serial Nos. 09/620,753, filed July 21, 2000; and 09/621,122, filed July 21, 2000.

CLAIMS

What is claimed is:

1. An isolated and purified polynucleotide sequence encoding an olfactory receptor and having the nucleotide sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO: 73 and SEQ ID NO:111 through SEQ ID NO:152, or a nucleotide sequence that is at least about 95% homologous to a nucleotide sequence of the group consisting of SEQ ID NO:1 through SEQ ID NO: 73 and SEQ ID NO:111 through SEQ ID NO:152 and encoding a polypeptide having olfactory receptor function.

2. An expression vector comprising a polynucleotide sequence of claim 1.

3. A host cell comprising the expression vector of claim 2.

4. An isolated and purified olfactory receptor polypeptide comprising the translated sequence of SEQ ID NO:1 through SEQ ID NO: 73 and SEQ ID NO:111 through SEQ ID NO:152, or a polypeptide sequence that is at least about 95% homologous to a polypeptide sequence of the group consisting of the translated sequence of SEQ ID NO:1 through SEQ ID NO: 73 and SEQ ID NO:111 through SEQ ID NO:152 and having olfactory receptor function.

5. A host cell expressing a polypeptide of claim 4 or a functional fragment thereof.

6. A phage expressing a polypeptide of claim 4 or a functional fragment thereof.

7. A preparation containing a polypeptide of claim 4, further comprising biological or synthetic molecules which maintain the functional structure of the polypeptide.

8. An isolated and purified polynucleotide sequence encoding an olfactory receptor and having the nucleotide sequence selected from the group consisting of SEQ ID NO: 153 through SEQ ID NO: 1084 or a nucleotide sequence having a sequence at least about 95% homologous to a nucleotide sequence of the group consisting of SEQ ID NO: 153 through SEQ ID NO: 1084 and encoding a polypeptide having olfactory receptor function.

9. An expression vector comprising a polynucleotide sequence of claim 8.

10. A host cell comprising the expression vector of claim 9.

11. An isolated and purified olfactory receptor polypeptide comprising the sequence of SEQ ID NO: 1085 through SEQ ID NO: 2008, or a polypeptide sequence that is at least about 95% homologous to a polypeptide sequence of the group consisting of SEQ ID NO: 1085 through SEQ ID NO: 2008 and having olfactory receptor function.

12. A host cell expressing a polypeptide of claim 11 or a functional fragment thereof.

13. A phage expressing a polypeptide of claim 11 or a functional fragment thereof.

14. A preparation containing a polypeptide of claim 11, further comprising biological or synthetic molecules which maintain the functional structure of the polypeptide.

15. A library of olfactory receptors suitable for determining the interaction pattern of a composition with the receptors, comprising the expression products of at least two polynucleotides of SEQ ID NO:1 through SEQ ID NO: 73, SEQ ID NO:111 through SEQ ID NO:152, and SEQ ID NO: 153 through SEQ ID NO: 1084 wherein said polynucleotides encode functional olfactory receptors; or functional fragments of said expression products.

16. A library of olfactory receptors according to claim 15, wherein the library comprises the expression products of at least 50 polynucleotides of SEQ ID NO:1 through SEQ ID NO: 73, SEQ ID NO:111 through SEQ ID NO:152, and SEQ ID NO: 153 through SEQ ID NO: 1084 wherein said polynucleotides encode functional olfactory receptors; or functional fragments of said expression products.

17. A library of olfactory receptors according to claim 15, wherein the library comprises the expression products of at least 100 polynucleotides of SEQ ID NO:1 through SEQ ID NO: 73, SEQ ID NO:111 through SEQ ID NO:152, and SEQ ID NO: 153 through SEQ ID NO: 1084 wherein said polynucleotides encode functional olfactory receptors; or functional fragments of said expression products.

18. A library of olfactory receptors according to claim 15, wherein the library comprises the expression products of at least 200 polynucleotides of SEQ ID NO:1 through SEQ ID NO: 73, SEQ ID NO:111 through SEQ ID NO:152, and SEQ ID NO: 153 through SEQ ID NO: 1084 wherein said polynucleotides encode functional olfactory receptors; or functional fragments of said expression products.

19. A library of olfactory receptors according to claim 15, wherein the library comprises the expression products of at least 500 polynucleotides of SEQ ID NO:1 through SEQ ID NO: 73, SEQ ID NO:111 through SEQ ID NO:152, and SEQ ID NO: 153 through SEQ ID NO: 1084 wherein said polynucleotides encode functional olfactory receptors; or functional fragments of said expression products.

20. A library of olfactory receptors suitable for determining the interaction pattern of a composition with the receptors, comprising at least two polypeptides of SEQ ID NO: 1085 through SEQ ID NO: 2008, wherein said polypeptides are functional olfactory receptors; or functional fragments of said polypeptides.

21. A library of olfactory receptors according to claim 20, wherein the library comprises at least 50 polypeptides of SEQ ID NO: 1085 through SEQ ID NO: 2008,

wherein said polypeptides are functional olfactory receptors; or functional fragments of said polypeptides.

22. A library of olfactory receptors according to claim 20, wherein the library comprises at least 100 polypeptides of SEQ ID NO: 1085 through SEQ ID NO: 2008, wherein said polypeptides are functional olfactory receptors; or functional fragments of said polypeptides.

23. A library of olfactory receptors according to claim 20, wherein the library comprises at least 200 polypeptides of SEQ ID NOS of SEQ ID NO: 1085 through SEQ ID NO: 2008, wherein said polypeptides are functional olfactory receptors; or functional fragments of said polypeptides.

24. A library of olfactory receptors according to claim 20, wherein the library comprises at least 500 polypeptides of SEQ ID NO: 1085 through SEQ ID NO: 2008, wherein said polypeptides are functional olfactory receptors; or functional fragments of said polypeptides.

25. A method for determining the binding pattern of a composition with olfactory receptors, comprising the steps of:

exposing the composition to a library according to claim 21; and

determining whether the composition binds to each olfactory receptor, thereby determining the overall binding patter of the composition.

26. The method of claim 25, wherein the composition consists essentially of one compound or chemical.

27. The method of claim 25, wherein the composition comprises at least two compounds or chemicals.

28. The method of claim 25, wherein the step of determining whether the composition binds to each olfactory receptor further comprises a determination of the

approximate binding constant with which the composition binds to each receptor or functional fragment thereof.

29. The method of claim 25, further comprising the step of determining whether a receptor or functional fragment thereof to which the composition binds is activated.

30. The method of claim 29, futher comprising the step of determining the absolute or relative amount by which the receptor or functional fragment thereof is activated.

31. A DNA array or a DNA chip comprising DNA segments derived from SEQ ID NO: 153 through SEQ ID NO: 1084.

32. A method of determining differences among individuals with respect to their olfactory faculties, comprising the steps of comparing the olfactory DNA of the individual against the array or chip of claim 31.

33. A method to determine single nucleotide polymorphisms in olfactory receptors, comprising the steps of uniquely amplifying olfactory receptor sequences from DNA obtained from one or more individuals, based on primers designed according to the first 25 bases and the last 25 bases of any combination of, or each of, SEQ ID NO: 153 through SEQ ID NO: 1084, and determining the similarities and differences between said amplified DNA and the corresponding receptor from SEQ ID NO: 153 through SEQ ID NO: 1084.

ctcattgttg gtactggttt gctcagatcc ttggatcacc tcccgatctt ggtggttggc 600
tgtgggggat tcaccctggt cacctctgtt gtggtgatcc ttgtctccta catgtcttcc 660
ctcatgacta tcctaggaat tcccttagct tctggaaaac agagagcctt ctccacctgt 720
gcctcccact tgactgctgt tagcctgtac tatgaaacaa ctatgtacac ttacttgccc 780
gcctcgcgac atggatccgg ggcaggaaat cagattgtgt cagtatttta tacaatggtg 840
atccccatgt taaatcctct catctatagt ttgagaaatg aggaagtgaa agttgcccta 900
tgaaaaacat tgagacatag tccttaatct tctattgagt gtctcaaaaa tgcaaaatat 960
tctgtgaaga 970

<210> 407
<211> 934
<212> DNA
<213> Unknown (H38g256 nucleotide)

<220>
<223> Synthetic construct

<400> 407
tggcttgatg aaaaaaaaca agattctaac gtgacagaac ttgttcttct gggcctatca 60
tcttcttggg agctgcagct atttctctta ttactatttt tgttttttta cattgctatt 120
gtcctgggaa acctcttgat agtggtaaca gtgcaagccc atgctcatct gctccaatct 180
cctatgtatt attttttagg tcatctctct ttcattgacc tatgcctaag ctgtgttact 240
ctgccaaaga tgttagggga tttcctacag cagggcaaga gcatctcttt ttcaggatgc 300
ctggccccaga tctacttcct ccactttcta ggagccagtg agatgttttt gctgacagtt 360
atggcctatg acaggtatgt tgccatctgt aaccctttgc gctaccttat aagtcatgaa 420
cccccagcta tgcctttggt tggttcttgc ctgctggtgt gggggtttta tccactctat 480
catgcaggtc atactagtca tccagctgcc tttctgtggc cccaatgaac tggacaactt 540
ctactgtgat gtcccacagg tcatcaagct ggcctgcatg gacacctatg tggtagaggt 600
gctgatgata gccaacagtg gtctgctctc tcttgtctgc ttcttggtct tactattctc 660
ttatgctgtc atcctgatca ccctgagaac acacttcggc cagggccaga acaagttcct 720
ctctacctgt gcttctcacc tgacagtggt cagcctgatc ttcatgccat gtatattcat 780
ctatttgagg cctttctgca gcttctctgt ggataagata ttctccatgt tttacacagt 840
gatgacacct atgttgagcc ccctcatcta cacactcaga aatgctgata tgaagacagc 900
tatgaagaag ctgaggataa aaccatgtga catt 934

<210> 408
<211> 954
<212> DNA
<213> Unknown (H38g257 nucleotide)

<220>
<223> Synthetic construct

<400> 408
atgatggtgg atcccaatgg caatgaatcc agtgctacat acttcatcct aataggcctc 60
cctggtttag aagaggctca gttctggttg gccttcccat tgtgctccct ctaccttatt 120
gctgtgctag gtaacttgac aatcatctac attgtgcgga ctgagcacag cctgcatgag 180
cccatgtata tatttctttg catgctttca ggcattgaca tcctcatctc cacctcatcc 240
atgcccaaaa tgctggccat cttctggttc aattccacta ccatccagtt tgatgcttgt 300
ctgctacaga tgtttgccat ccactcctta tctggcatgg aatccacagt gctgctggcc 360
atggcttttg accgctatgt ggccatctgt cacccactgc gccatgccac agtacttacg 420
ttgcctcgtg tcaccaaaat tggtgtggct gctgtggtgc ggggggctgc actgatggca 480
ccccttcctg tcttcatcaa gcagctgccc ttctgccgct ccaatatcct ttcccattcc 540
tactgcctac accaagatgt catgaagctg gcctgtgatg atatccgggt caatgtcgtc 600
tatggcctta tcgtcatcat ctccgccatt ggcctggact cacttctcat ctccttctca 660
tatctgctta ttcttaagac tgtgttgggc ttgacacgtg aagcccaggc caaggcattt 720
ggcacttgcg tctctcatgt gtgtgctgtg ttcatattct atgtaccttt cattggattg 780
tccatggtgc atcgctttag caagcggcgt gactctccgc tgcccgtcat cttggccaat 840
atctatctgc tggttcctcc tgtgctcaac ccaattgtct atggagtgaa gacaaaggag 900
attcgacagc gcatccttcg acttttccat gtggccacac acgcttcaga gccc 954

145 150 155 160
Met Gln Val Ile Leu Val Ile Gln Leu Pro Phe Cys Gly Pro Asn Glu
165 170 175
Leu Asp Asn Phe Tyr Cys Asp Val Pro Gln Val Ile Lys Leu Ala Cys
180 185 190
Met Asp Thr Tyr Val Val Glu Val Leu Met Ile Ala Asn Ser Gly Leu
195 200 205
Leu Ser Leu Val Cys Phe Leu Val Leu Leu Phe Ser Tyr Ala Val Ile
210 215 220
Leu Ile Thr Leu Arg Thr His Phe Gly Gln Gly Gln Asn Lys Phe Leu
225 230 235 240
Ser Thr Cys Ala Ser His Leu Thr Val Val Ser Leu Ile Phe Met Pro
245 250 255
Cys Ile Phe Ile Tyr Leu Arg Pro Phe Cys Ser Phe Ser Val Asp Lys
260 265 270
Ile Phe Ser Met Phe Tyr Thr Val Met Thr Pro Met Leu Ser Pro Leu
275 280 285
Ile Tyr Thr Leu Arg Asn Ala Asp Met Lys Thr Ala Met Lys Lys Leu
290 295 300
Arg Ile Lys Pro Cys Asp Ile
305 310

<210> 1340
<211> 318
<212> PRT
<213> Unknown (H38g257 protein)

<220>
<223> Synthetic construct

<400> 1340
Met Met Val Asp Pro Asn Gly Asn Glu Ser Ser Ala Thr Tyr Phe Ile
1 5 10 15
Leu Ile Gly Leu Pro Gly Leu Glu Glu Ala Gln Phe Trp Leu Ala Phe
20 25 30
Pro Leu Cys Ser Leu Tyr Leu Ile Ala Val Leu Gly Asn Leu Thr Ile
35 40 45
Ile Tyr Ile Val Arg Thr Glu His Ser Leu His Glu Pro Met Tyr Ile
50 55 60
Phe Leu Cys Met Leu Ser Gly Ile Asp Ile Leu Ile Ser Thr Ser Ser
65 70 75 80
Met Pro Lys Met Leu Ala Ile Phe Trp Phe Asn Ser Thr Thr Ile Gln
85 90 95
Phe Asp Ala Cys Leu Leu Gln Met Phe Ala Ile His Ser Leu Ser Gly
100 105 110
Met Glu Ser Thr Val Leu Leu Ala Met Ala Phe Asp Arg Tyr Val Ala
115 120 125
Ile Cys His Pro Leu Arg His Ala Thr Val Leu Thr Leu Pro Arg Val
130 135 140
Thr Lys Ile Gly Val Ala Ala Val Val Arg Gly Ala Ala Leu Met Ala
145 150 155 160
Pro Leu Pro Val Phe Ile Lys Gln Leu Pro Phe Cys Arg Ser Asn Ile
165 170 175
Leu Ser His Ser Tyr Cys Leu His Gln Asp Val Met Lys Leu Ala Cys
180 185 190
Asp Asp Ile Arg Val Asn Val Val Tyr Gly Leu Ile Val Ile Ile Ser
195 200 205
Ala Ile Gly Leu Asp Ser Leu Leu Ile Ser Phe Ser Tyr Leu Leu Ile
210 215 220
Leu Lys Thr Val Leu Gly Leu Thr Arg Glu Ala Gln Ala Lys Ala Phe
225 230 235 240

Gly Thr Cys Val Ser His Val Cys Ala Val Phe Ile Phe Tyr Val Pro
245 250 255
Phe Ile Gly Leu Ser Met Val His Arg Phe Ser Lys Arg Arg Asp Ser
260 265 270
Pro Leu Pro Val Ile Leu Ala Asn Ile Tyr Leu Leu Val Pro Pro Val
275 280 285
Leu Asn Pro Ile Val Tyr Gly Val Lys Thr Lys Glu Ile Arg Gln Arg
290 295 300
Ile Leu Arg Leu Phe His Val Ala Thr His Ala Ser Glu Pro
305 310 315

<210> 1341
<211> 320
<212> PRT
<213> Unknown (H38g258 protein)

<220>
<223> Synthetic construct

<221> VARIANT
<222> (1)...(320)
<223> Xaa = Any Amino Acid

<400> 1341
Met Ser Ser Arg Leu Met Asn Val Phe Ser Met Glu Thr Ile Asn Phe
1 5 10 15
Val Ser Cys Leu Ile Leu Met Gly Phe Pro Ser Ser Pro Glu Met Gln
20 25 30
Leu Leu Tyr Phe Gly Leu Phe Ser Val Ala Tyr Thr Leu Thr Pro Met
35 40 45
Gly Asn Ala Ala Ile Val Cys Ala Val Trp Xaa Asp Gln His Leu His
50 55 60
Thr Pro Met Tyr Thr Leu Leu Gly Asn Phe Ser Leu Leu Glu Ile Trp
65 70 75 80
Tyr Val Thr Ala Thr Lys Leu Leu Ala Asn Phe Leu Ser Thr Ser Lys
85 90 95
Ser Ile Ser Phe Met Ser Cys Phe Ala Gln Phe Tyr Phe Phe Ser Leu
100 105 110
Gly Tyr Asp Glu Gly Phe Phe Leu Cys Ile Thr Ala Phe Asp Arg Tyr
115 120 125
Leu Ala Ile Cys Arg Pro Leu Arg Tyr Pro Cys Ile Met Thr Lys Gln
130 135 140
Val Cys Thr Gly Leu Ile Ile Phe Ala Trp Ser Cys Val Phe Val Ile
145 150 155 160
Phe Leu Thr Leu Val Ile Leu Ile Ser Gln Leu Ser Tyr Cys Gly Pro
165 170 175
Asn Ile Ile Asn His Phe Ile Cys Asp Pro Val Pro Leu Lys Met Leu
180 185 190
Ser Cys Ser Glu Asp Ile Ile Ile Thr Gln Leu Ile Tyr Ser Thr Phe
195 200 205
Asn Ser Val Phe Ile Ile Gly Thr Phe Leu Phe Ile Leu Cys Ser Tyr
210 215 220
Ala Leu Val Ile Leu Ala Ile Ile Arg Met Pro Ser Glu Ala Gly Lys
225 230 235 240
Arg Lys Ala Phe Ser Thr Cys Ala Ser His Leu Ala Val Val Thr Leu
245 250 255
Phe Tyr Gly Ser Ile Met Val Met Tyr Val Ser Pro Gly Ser Ala His
260 265 270
Pro Ala Lys Asn Glu Lys Ile Ile Thr Leu Phe Phe Ser Val Ile Thr
275 280 285
Pro Leu Cys Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Glu Met Lys

FIGURE 1

SEQ. ID NO:1
```
  1 GGNTTATNCC NCGTTGNACT GCAGGGGNNC AACNCACAGN ACGCCCGNTG CTGAGGCTAT AAATGANCGG
 71 NTTAAGGAGA GGAGTGAAGA CAGTAAAAAA ACACAGAGAT AAATTTATCA ATTGGGAAGC TTTCAAAGGG
141 CCAAATATAG ATGAATATTA ATGGGCCAAA GAAGAGAAGC ACAACAGTAA TGTGGGCAGA CAGAGTGGAA
211 AGGGCCTTGG ACATCCCATC AGAGGCTTGG CGATGCACAG TAGCAAGGAT GATAGTGTCA GAAATGAGCA
281 AAAGGAGGAA ACACATAAGT GAGAGCAGAC CACTGTTAGT GAGCACCAGT ATCTCAAAAC CATAGGTGTC
351 TAAGCAGGCA AGCTTGATCA CTAGGAGGAG GTCACAGAAA AAATTGTCTA CCCTGTTGGG TCCACAGAAA
421 GGCAGATTGA CTTTGAATGC CAGGTGGGTG GCTGAGTGTG AGATGCCAAT GGCCCAGGAA ACCCCCACCA
491 GAACAGTTCA CACCCTCCGG TTCATGATGG TTATGTAGTG CAGAGGTTTG CATATAGCAA TGTATCTATC
561 ATAGGCCATG GCAACAAGAA GCACCATCTC ACTACCCCCA AAAACATGCA AGN
```

SEQ. ID NO:2
```
  1 GGNNTNTNAC ACGGACTCCA AGCAGTGGTA ACAACGCAGA GTACGCCCGT TCCTGAGTGA GTAGATGAAG
 71 GGGTTCAGCA TGGGATTGAT GACAGTGTTG AAAATTCCAA CAGCTTTATC CTTGTCTGAA AGCTTGGTTG
141 AACCCAGTCG CATATAGTTA AAGATACCTG AACCATAGAA TATGGCAACC ACAGTGAGGT GGGAGCCACA
211 TGTGGAGAAG GCTTTCTTCC TGCCCTCTAC AGAGCGAATT CGCAGGACTG CAGCTGCCAC GTGGATATAG
281 GAGATGACAA TGAGAGCCAT GGGGGTACCT GCCATTATAA AACCCACAGC AAAAAGCAGC AGCTCATTGA
351 GTTGGGTGCT GGAGCAGGAG AGCTGGAAGA GCTGTGGGAG GTCACAGTAG AAGTGATTGA TCACATTGGG
421 GCCACAGAAG TTGAGCGTGG ACATGGCCAC AGTGTGGGTC AGTGCGTTGG TGAAAGCACA AGCCCAGGAC
491 GCAGCCACCA ACATCCTCTG GACTGTCTGA CTCATGCGGG TGCTTGTAGG TGAGGGGCCC GGCAGATGGG
561 CAGGAATCGG TCATAGGG
```

SEQ. ID NO:3
```
  1 TGGNNTTTTA TCNCCNTTGG AGCTCCNAAG CAGTGGTAAC AACGCAGAGT ACGCCCGTTG CGAAGCGTGT
 71 AGATTAGGGG GTTCAGTAGG GGAGTGATGA CAGTGTAGGT CACCGAGATC AGCTGGTCAT GTTCTCTGGT
141 GTTCTCTGAC TTGGGCTTGA GGTAGGCAAT GGAGGCACAG CTGTAGTGGA CAATGACCAC AGTGAGGTGG
211 GATGCACAGG TGGCAAAAGC CTTCTTCCGG CCCTCAACTG AAGTAATCTT GAGGATTGTA GAGATAATGA
281 GAACATAAGA AATGAAAACC AGACCCATAG GTACAACAAG CACCAGCACA CTGATAATCA AAGTCAGGAT
351 TTCATTGACA GTGGTGTCAA TGCAGGAGAG CTTCATCACA GGGCGGATGT CACAGAAGAA GTGGGGCACC
421 TTTTCTAGCA CAGAAGGGTA ACCTGAATAC AGATGTCACT TGCGTTATTG CTACAATCAG CCCAATGCTG
491 CAAGGCCCCC AGGACAAGTT GGATACGCAG CCTCTTGTTC ATAATAACCA TGTATCTCAA GGGGGTTGCA
561 AGATGGCCAC ATAGCNGNTC ATATTCCN
```

SEQ. ID NO:4
```
  1 GTNGTTNTTA ACNCCATTGG AGCTCCAAAG CAGTGGTAAC AACGCAGAGT ACGCCCCCAA TGTATTTTTT
 71 TTTGAGAAAC TTGTCTTTCT TAGATTTTTG TTACATCTCT GTCACAATTC CAAAATCTAT TGTTAGTTCC
141 TTGACTCATG ATACTTCCAT TTCTTTCTTT GGGTGTGCTC TGCAAGCCTT CTTTTTCATG GACTTGGCAA
211 CTACGGAGGT AGCCATCCTT ACAGTGATGT CCTGTGACCG CTATATGGCC ATCTGCCGGC CTTTACATTA
281 TGAGGTCATC ATAAACCAAG GTGTCTGTCT GAGGATGATG GCCATGTCGT GGCTCAGTGG GGTGATCTGT
351 GGATTCATGC ATGTGATAGC AACATTCTCA TTACCATTCT GTGGGCGCAA TAGAATACGT CAATTTTTCT
421 GTAATATTCC ACAACTNCTA AGCCTCTTAG ACCCCAAAGT AATTACCATT GAGATTGGAG TCATNGGNTT
491 TTGGTACAAG TCTTGNGATA ATCCTCTTTG NTGNAATTAC TCTCTCCTAC ATGTNCATTT TTTTTTGNCA
561 TCATGAGGGA TTCCTTCTAA AGG
```

SEQ. ID NO:5
```
  1 GNGGNTTNNT NCCNCCNTTG GACTCCAAAG CAGTGGTAAC AACGCAGAGT ACGCCCGTGT GTAAATGAAT
 71 GGGTTCAACA TGGGAGTCAT AACAGTGTAG GACAATGATA GCAGCTTCGT GCCCTCAGGT GAATTATTTG
141 ATTTAGGCCG GAAGTAGGTG AGGCTTAATG ATATATAGAA AAGAGAGACA ACAAGGAGGT GTGAGGAACA
211 TGTAGAAAAG GCTTTATTCT TCCCTTTAGC TGATGGGATC TTGAGGATGG CAGCAGCAAT GCGAGTATAG
281 GAACACAAGA TCAGCAAGCA GGGGATCATG ACCACCAGAA TGGTTCCGAC GATGGCGTAG ATCTCAAACA
351 GTGCTGTGTC TGCACAGACC AGCCTCAGCA CAGGTGGGCT GTCACAGAAG AAGTGGTTCA CCTTGTTGGT
421 GCCACAGAAT GGAAAACTGA AGAGCCATGT GGTCTGCACA GTAGCTACAG GAAAGCCTGG GAACCAGGAG
491 GCAGCAGCCA GTTTGGCACG AGTCCTTTGG TTCATGATGA CTGGGTAGTG CAAGGGACTN GCAGATNNNC
```

561 NCATTCGGTC ATATGNCATG GNAG

SEQ. ID NO:6
1 CNTTGGAGCT CCAAAGCAGT GGTAACAACG CAGAGTACGC CCGCTCCGCA GAGAATAGAT GAAAGGGTTC
71 AGGGTCGGGG GCACGACTGT GTAGAACGCA GACAGGAAAA CATCCAGAAC GGGGGGAGAA TTTGAAATTG
141 GCTTCACATA GGCAATGCTG CCAGATATCA TAAAGAGTGT TACAACCACA AGATGTGGAA TGCAGGTAGA
211 AAATGTTTTT GATCTACCCT CCTTAGAAGG AATCCTCATG ATGACAGAAA AAATGTACAT GTAGGAGAGA
281 GTAATTACAA CAAAGGAGAT TATCACAAGA CTTGTACCAA AAACCATGAC TCCAATCTCA ATGGTAATTA
351 CTTTGGGGTC TAAGAGGCTT AGGAGTTTGT GGAATATTAC AGAAAAATTG ACGTATTCTA TTGCGCCCAC
421 AGAATGGTAA TGAGAATGTT GCTATCACAT GCATGAATCC ACAGATCACC CCACTGAGCC ACGACATGGC
491 CATCATCCTC AGACAGACAC CTTGGTTTAT GATGACCTCA TAATGTAAAG GCCGGCAGGA TGGCCATATA
561 GCGGTCATAG GA

SEQ. ID NO:7
1 GCAGTGGTAA CAACGCAGAG TACCGCCCCC TATGTACTTT TTCTTGGGAA ACTTGTCTGT GTTTGACATG
71 GGTTTCTCCT CAGTGACTTG TCCCAAAATG CTGCTCTACC TTATGGGGCT GGGCCGACTC ATCTCCTACA
141 AAGACTGTGT CTGCCAGCTT TTCTTCTTCC ATTTCCTCGG GAGCATTGAG TGCTTCTTGT TTACGGTGAT
211 GGCCTATGAC CGCTTCACTG CCATCTGTTA TCCTCTGCGA TACACAGTCA TCATGAACCC AAGGATCTGT
281 GTGGCCCTGG CTGTGGGCAC ATGGCTGTTA GGGTGCATTC ATTCCAGTAT CTTGACCTCC CTCACCTTCA
351 CCTTGCCACA CTGTGGTCCC AATGAAGTGG ATCACTTCTT CTGTGACATT CCAGCACTGT TGCCCTTGGC
421 CTGTGCTGAC ACATCCTTAG CCCAGAGGGT GAGCTTCACC AACGTTGGCC TCATATCTCT GGCTGCTTTC
491 TGCTAAATCT TTTATCCTAC ACTAGAATCA CAAATATCTA TCTTAAGCAT TCGTACAAC

SEQ. ID NO:8
1 GGAACAACGC AGAGTCGCCC CCGATGTACT TGTTCTTCTC CAACCTGTCC TTTGCTGACA TTTGTGTTAC
71 TTCCACCACC ATTCCAAAAA TGCTGATGAA CATCCAGACA CAGAACAAAG TCATCACCTA CATAGCCTGC
141 CTCATGCAGA TGTATTTTTT CATACTCTTT GCTGGATTTG AAAACTTCCT CCTGTCCGTG ATGGCCTATG
211 ACCGGTTTGT GGCCATCTGT CACCCCCTGC ACTACATGGT CATTATGAAC CCTCACCTCT GTGGACTGCT
281 GGTTCTGGCA TCCTGGACCA TGAGTGCTCT GTATTCCTTG CTACAAATCT TAATGGTAGT ACGACTGTCC
351 TTCTGCACAG CCTTAGAAAT CCCCCACTTT TTCTGTGAAC TTAATCAGGT CATCCAACTT GCTTGTTCTG
421 ATAGCTTTCT TAATCACATG GTGATATATT TTACAGTTTG CGCTGCTGGG TGGAGGTCCC TGACTGGGAT
491 CCTTTACTTC TTACTCTAAG ATAATTTCTT CATACATGCA ATCTCANCAA GNTCAGGG

SEQ. ID NO:9
1 GGGTTTTNAC CCNNTNGGAG CTCCNAGCAG TGGTAACAAC GCAGAGTACG CCCGTTTCGT AGGCTATAAA
71 TGAAGGGGTT GAGTGAGGGA GTCACCACTC CATAGAAGAG GGCCATGAAC TTGGGTTGAT CCCTTGAGAT
141 GGAGGAGGGG GGCTGAAGGT ACATGCTGAT GGCTGGGCCA TAAAATAAGA AAACTACAAT AAGATGGGAG
211 GAGCATGTCC CAAAGGCCTT TNTCCTTCCC TTGGAAGATT TGATCTTAAA TACAGCACTT NCAATACTAG
281 CATAGGAAGC AAGAATTAAG CATANTGGGA CAGCTAACAT AAAAATGCAT ACCACAGAGA GTGTGAGCTC
351 GTTAGAACCC TTTTCACCAC AGGCAATCTT TATCAGAACA GGAATCTCAC ACACCAAGTG GTCCAGCTTA
421 TTGAGACCAC ACAGTGGNAA TTTGTATTGT GGCAGTGGCC CTCTGAGAAC GGCATAGATT ATACCAANTT
491 AACCACNACN GCGGNAACTA ANGATTCAGA CGCNCTGGAT TCATGATGAG GGTNTAGTGA AGAGGTTNTC
561 AGAATGGCCA CATACCGNTC AAA

SEQ. ID NO:10
1 GCTGCTNCCA GCAGTGGTAA CAACGCANAG TACGCCCCCA ATGTATTTGT TCTTCGGCCA TCTGTCTCTC
71 CTGGATGTCT GCTTCATCAC CACTACCATC CCACAGATGT TGATCCACCT CGTGGTCAGG GACCACATTG
141 TCTCCTTTGT ATGTTGCATG ACCCAGATGT ACTNTGTCTT CTGTGTTGGT GTGGCCGAGA GCATCCTCTT
211 GGCTTTCATG GCCTATGACC GNTATGNTGC TATCTGCTAC CCACTTAACT ATGTCCCGAT CATAAGCCAT
281 AAGGTCTGTG TCAGGCTTGT GGGAACTGCC TGGNTCTTTG GGCTGATCAA TGGCATCTTT NTCGGGTATA
351 TTTCATTCCT AGAGCCCTTC CGCAGAGACA ACCACATAGA AAGCTTCTTC TGCGAGGCCC CCATAGTGAT
421 TTGGCCTCTT TTGTGGGGGA CCCTNANANT AGTCTGTGGG CAAATCTTTN GCCGATGCCA TCGTGGTAAT
491 TCTNAGNCCC ATNGGTGCTN ACTGNTACTT ACCTATNTGC ACATTCCTGT CCACCATCCT AGNNAAAGTC
561 CTCCTTCTN

SEQ. ID NO:11

1 GGNNTTTTAC CNCNATTGGA GCTCCAAAGC AGTGGTAACA ACGCAGAGTA CGCCCCCTAT GTACTTGTTC
71 TTGAGAAACT TGTCTTTCTT AGATTTTTGT TACATCTCTG TCACAATTCC AAAATCTATT GTTAGTTCCT
141 TGACTCATGA TACTTCCATT TCTTTCTTTG GGTGTGCTCT GCAAGCCTTC TTTTTCATGG ACTTGGCAAC
211 TACGGAGGTA GCCATCCTTA CAGTGATGTC CTATGACCGC TATATGGCCA TCTGCCGGCC TTTACATTAT
281 GAGGTCATCA TAAGCCAAGG TGTCTGTCTG AGGATGATGG CCATGTCGTG GCTCAGTGGG GTGATCTGTG
351 GATTCATGCA TGTGATAGCA ACATTCTCAT TACCATTCTG TGGGCGCAAT AGAATACGTC AATTTTTCTG
421 TAATATTCCA CAGCTCCTAA GCCTCTTAGA CCCCAAAGTA ATTACCATTG AGATTGGAGT CATGGTTTTT
491 GGTACAAGGC TTGNGATAAT CTNCTTTGGT GNAATTACTC TCTCCTACAT GTACATTTTT TCTGCATCAT
561 GAGGATTCCT TCTAAGGAGG GG

SEQ. ID NO:12

1 GGNTTTGACC ACGGAGCTCC AAGCAGTGGT AACAACGCAG AGTACGCCCT CTTGTCCTCG TGCCGATACA
71 TGATGGGGTT CAACATGGGA GTCATAACAG TGTAGGACAA TGATAGCAGC TTCTTGCCCT CAGGTGAATT
141 ATTTGATTTA GGCCGGAAGT AGGTGAGGCT TAATGATATA TAGAAAAGAG AGACAACAAG GAGGTGTGAG
211 GAACATGTAG AAAAGGCTTT ATTCTTCCCT TTAGCTGATG GGATCTTGAG GATGGCAGCA GCAATGTGAG
281 TATAGGAACA CAAGATCAGC AAGCAGGGGA TCATGACCAC CAGAATGGTT CCGACGATGG CGTAGATCTC
351 AAAGAGTGCT GTGTCTGCAC AGACCAGCCT CAGNACAGGT GGGCTGTCAC AGAAGAAGTG GTTCACCTTG
421 TTGGTGCCAC AGAATGGAAA ACTGAAGAGC CATGTGGTCT GCACAGTAGC TACAGGAAAG CCTGGGAACC
491 AGGAGGTAGC AGCCAGTTTG CACGAGTCCC TTTGGTTNAT GAATGACTGG GGTAGTGCAA GGGACTGCAG
561 ATGGCCACAT ANCGGTCNT

SEQ. ID NO:13

1 GNNNTTNNNN CCACTGGAGC TCCAAAGCAG TGGTAACAAC GCAGAGTACG CCCCCAATGT ATTTATTCTT
71 GCTCACCTCT CCTTAGTTGA TATCTGTTTT ACCACCAGTA TTGTCCCCCA GCTGCTGTGG AACCTAAAAG
141 GACCTGACAA AACAATCACA TTCCTGGGTT GTGTCATCCA GCTCTACATC TCCCTGGCAT TGGGCTCCAC
211 TGAGTGTGTC CTCCTGGCTG TAATGGCTTT TGATCGCTAT GCTGCAGTTT GCAAACCTCT CCACTATACC
281 GCCGTAATGA ACCCTCAGCT GTGCCAGGCT CTGGCAGGGG TTGCGTGGCT GAGTGGAGTG GGAAACACTC
351 TTATCCAGGG CACTGTCACC CTCTGGCTTC CTCGCTGTGG ACACCGATTG CACTAACATT TCTTCGTGAG
421 GTACCCTCCA TGATTAAGCT TGCATGTGTG GACATCCATG ATAATGAGGT TCAGCTCTTT GTTGCTTCAC
491 TGGTCTTGCT CCTCTTGCCC TTAGTGCTAA TACTGCTGCC TATGGACATA TAGCCAAGGT GGCATAAGGA
561 TCAAGTCAGT CCAGCCT

SEQ. ID NO:14

1 GGNNTNTNAC TCCATGGACT CCAAGCAGTG GTAACAACGC AGAGTACGCC CATACATGAT GGGGTTCAGT
71 AGGGGAGTGA TGACAGTGTA GGTCACCGAG ATCAGCTGGT CATGTTCTCT GGTGTTCTCT GACTTGGGCT
141 TGAGGTAGGC AATGGAGGCA CAGCTGTAGT GGACAATGAC CACAGTGAGG TGGGATGCAC AGGTGGCAAA
211 AGCCTTCTTC CGGCCCTCAA CTGAAGCAAT CTTGAGGATT GNAGAGATAA TGAGAACATA AGAAATGAAA
281 ACCAGACCCA TAGGTACAAC AAGCACCAGC ACACTGATAA TCAAAGTCAG GATTTCATTG ACAGTGGTGT
351 CAATGCAGGA GAGCTTCATC ACAGNGCGGA TGTCACAGAA GAAGTGGGGC ACCTTTCTAG CACAGAAGGG
421 TAACCTGAAT ACAGATGTCA CTTGCGTTAT TGCTACAATC AGCCCAATGC TGCNGGCCCC CAGGACAAGT
491 TGGATACGCA GCCTTNTCGT TCTANTAACC ATGTATCTCA ANGGGCTTGC NGATNNCCAC ATACTNGCAT
561 ANACCATTGC TGNGAGC

SEQ. ID NO:15

1 GNCGNTNTTA ACNCCATTGG AGCTCCAAAG CAGTGGTAAC AACGCAGAGT ACGCCCATTA CGAAAAGTGT
71 AGATGAAGGG GTTCAAGAGG GGTGTGATGA TGCAGCTCAG GACGGAGGCA CCTTTGTTGA GCAGTTTGGA
141 CTGAGCCTCT GACATACGAA TGTAGAGAAA GATGGAACTG CCATAGATGA TGACCACCAC TGTAAGATGC
211 GAGGCGCAAG TGGAAAACGC TTTCCTTCGC TCAGCAGCTG TAGGGGCCCT GAGAACAGTG GCAAGAATGC
281 AGGCATAGGA AACTGAGGTC AGAGCCAGTG AGCCCAGTAA CACCAACGTA GAGAGCATGA AAGCCACCAG
351 TTTCAGCAGG TGGGTGTCCC CACAAGAAAG CCTGAGCAAG GGCCAACTGT CACGAAAGAA GTGGTCAATA
421 CCATTGNGGC CACAGAAAGG CATGGCTGGC CATGAGGACA GTGGGGCAAA GGACCCAGAG GAATNCANCT
491 AGCCAGGAGG CCACACTAGT TTGTGAACAG ACATGGCCAT TNATTAGGGT CTCATAGCGG AGTTGTCGNC
561 AGATTTGCNT GGTNACGATT CAN

SEQ. ID NO:16

1 GGNNTTTTAC CNCNATTGGA CTCCAAAGCA GTGGTAACAA CGCAGAGTAC GCCCCCTATG TATTTATTCT
71 TGCTCACCTC TCCTTAGTTG ATATCTGTTT TACCACCAGT ATTGTCCCCC AGCTGCTGTG GAACCTAAAA
141 GGACCTGACA AAACAATCAC ATTCCTGGGT TGTGTCATCC AGCTCTACAT CTCCCTGGCA TTGGGCTCCA
211 CTGAGTGTGT CCTCCTGGCT GTAATGGCTT TTGATCGCTG TGCTGCAGTT TGCAAACCTC TCCACTATAC
281 CGCCGTAATG AACCCTCAGC TGTGCCAGGC TCTGGCAGGG GTTGCGTGGC TGAGTGGAGT GGGAAACACT
351 CTTATCCAGG GCACTGTCAC CCTCTGGCTT CCCCGCTGTG GACACCGATT GCTCCAACAT TTCTTCGTGA
421 GGTACCCTCC ATGATTAAGC TTGCATGTGT GGACATCCAT GATAATGAGG TTCAGCTCTT TGTTGCTTCA
491 CTGGTCTTGC TCCTCTTGCC CTTAGTGCTA ATACTGCTGC CTATGGACAT ATAGCCAANG TGGCATAAAG
561 GATCAAGTCA GTCCAGG

SEQ. ID NO:17

1 GNNNNTTNTT CANTCCATTG GGCCCTCTAG ATGCATGCTC GAGCGGCCGC CAGTGTGATG GATATCTGCA
71 GAATTCGCCC TTATTCCGGA GGGTATACAT GAAGGGATTG GTAACTAGAC GTAAACTCGA AGCCAAGAAC
141 AGAATTTCTC TTAGAAAAGA GAATTGAAAC TAAAGAGAAA GAACTAGCAA AGAAGGAAAT ATTGAATATA
211 CAAGAGAGAG GAGACAGATG ATGGAACAAG ACTCTGAAAG AGGTGGAAGG GATTGAATAC AATCAAAAGT
281 ATGGTGACTG CTAGTTCCAA GATGGTGGCG TAGGGGCAAG CTGGCTTTGC TTACCCCCCT GGCAGAAAAC
351 CAAAAACAAA TAGCACCAAG ATTATCACTA GCAATATCCC AGAACTCACA TATAAGGATG AGACAGTTCC
421 CAGGGCCCAG AGAAGATCAG AAGCACAAGT GGGAGAAGTC AGCTTTGGAT GCTACTTTGT TCTAAGGGAG
491 ACAAGTTGGG AGGATGATTG CAGATGTATA TTCAATGTTA TAAAACAGCC CATAAAACAA AGATTGGAAA
561 ATGTTGAATT TTGCAACCAG GAGCAAATAC TGGGAAAGGC GAATTCCAGC CACTTGCNGC C

SEQ. ID NO:18

1 GNNNNTTNAN TCANTGCCCT NGGGCCCTCT AGATGCATGC TCGAGCGGCC GCCAGTGTGA TGGATATCTG
71 CAGAATTCGC CCTTGTTGCG CAAGGTGTAA ATGAAAGGGT TTGCGCAGGA GTAAATGAAG GGATTACGCA
141 GGAGTAAATG AAGGGATTAC GCAGGAGTAA ATGAAGGGAT TACGCAGGAG TAAATGAAGG GATTACGCAG
211 GAGTAAATGA AGGGATTACG CAGGAGTAAA TGAAGGGATT ACGCAGGAGT AAATGAAGGG ATTACGCAGG
281 AGTAAATGAA GGGATTACGC AGGAGTAAAT GAAGGGATTA CGCAGGAGTA AATGAAGGGA TTACGCAGGA
351 GTAAATGAAG GGATTACGCA GGAGCAAATA CATAGGAAGG GCGAATTCCA GCACACTGGC GGCCGTTACT
421 AGTGGATCCG AGCTCGGTAC CAAGCTTGAT GCATAGCTTG AGTATTCTAA CGCGTCACCT AAATAGCTTG
491 GCGTAATCAT GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG
561 CCCGGAAGCA TAAAGTGTAA AGNCTGGGGT GCCTAATGAG TGACTTACTC CATTA

SEQ. ID NO:19

1 GNNANTNATT CCATCCATTG TCCCTTCAGA TGCATGCTCG AGCGGCCGCC AGTGTGATGG ATATCTGCAG
71 AATTCGCCCT TCTTGGTTTT TGTGCTGATA GATCATGGGA TTCAGCATGG GGGTGACCAC AGTGTACATC
141 ACTGAGGCTG TTGCACTTGA GTGTGAGTTG CGGGTGGCAG CAGAACTAAG GTACACCCCT AGGATTGCAC
211 CATAAAATAA GGAGACAACT GAGAGGTGAG ATGCACAGGT GGAAGATGCC TTGTACTTCC CCTGAGCTGA
281 TGAGATNGCA TGTATGGAAN GAAATTATNT TANAAGTAAG AGTAAAGNAT NCCAGTCAGG GGNANCNTTC
351 ACCCATCAGN TGCAANTTGT AAAAATTATA TTCAANCNAT NTGNATTTAA NGAAAANCCT TATCANGTAN
421 ACACTGCNAA GNTNTGNATT NANCCCTNGN ANTTAANNTT TCNACAAGAA AATAANGTGC GTTNNAATCT
491 TTNTAAGTCC CTNTCNCCAT TAANGTCNAN TCCNTCCNTA TCCCTTTTCN NATTTTGNAN TCNNGANTAC
561 NNTCTNNNGC NNTCNATTTC TNTNNTNNCT GACCTACTAA CCNATTNAGT TACNACAAGN CCNTTCNANT
631 CTCTATAATT NCTCGCANGT TNTCCCTCTT NNCANNTNCC CNTTNTTNTC CCTNTTCCCC ATCTNC

SEQ. ID NO:20

1 CCATTGGCCC TCTAGATGCA TGCTCGAGCG GCCGCCAGTG TGATGGATAT CTGCAGAATT CGCCCTTCCT
71 ATGTATTTTC TCTTACTGGG CTTTCCTGGT TCTCAAACTC TTCAGCTCTC TCTCTTTATG CTTTTTCTGG
141 TGATGTACAT CCTCACAGTT AGTGGTAATG TGGCTATCTT GATGTTGGTG AGCACCTCCC ATCAGTTGCA
211 TACCCCCATG TACTTCTTTC TGAGCAACCT CTCCTTCCTG GAGATTTGGT ATACCACAGC AGCAGTGCCC
281 AAAGCACTGG CCATCCTACT GGAGAGAAGT CAGACCATAT CATTTACAAG CTGTCTTTTG CAGATGTACT
351 TTGTTTTCTC ATTAGGCTGC ACAGAGTACT TCCTCCTGGC AGCCATGGCT TATGACCGCT GTCTTGCCAT
421 CTGCTATCCT TTACACTACG GAGCCATCAT GAGTAGCCTG CTCTCAGCGC AACTGGCCTT GGGCTTCTGG
491 GTGGNTGGGT TCGGGGGCAA TGCAGTGCCC ACAGGCCTTC AATCAAGTGG GCTGNTCCTT CTGGTGGCCC
561 CCGGTGCCAA TCAACCACTT TNTTTTTGGG ACAATTGCAN CCCTGGAATT GGCCC

SEQ. ID NO:21

1 GNNCTTANTT CAATCCCACC NANCCNTGCC GANGCATGCT CGNGCGGCCG CCAGTGTGAT GGATATCTGC
71 AGAATTCGCC CTTCCTATGT ATTTACTCTT ACTGGGCTTT CCTGGNTCTC AAACTCTTCA GCTCTCTCTC
141 TTTATGCTTT TTCTGGTGAT GTACATCCTC ACAGTTAGTG GTAATGTGGC TATCTTGATG NTGGTGAGCA
211 CNTCCCATCA GNTGCATACC CCCATGTTNT TCTTTCTGAG CNACCTCTCC TTCCTGGAGA TTTGGTATNC
281 CNCAAGCNGC ANNGCCCAAA GCTTTGCNCA TCTTATTGCN CAGANGCNNN CCNNTACANN NACNCTCCTG
351 TTTNTCGCTN CCTTNCCTCT TNCTTCNCTC ANNTACTNCN TCTNCTNTAG TNTCTTTCTT CTCTNTCNCT
421 CNTNNCNCCT NTAATNTTCC NCCTNTTCTN NTTTCTNTTT TCCCTNCTCT GTTTCACCCC TACCTCTTAT
491 CCNTNCTNCT NACTTCANNC TCNGNCNNTN NNNCNCCNNT AAATNTANGN NNANNNTNNN ATNTNCTCTT
561 CTCCNTTTAT ATCGCCTCTT CTCNTNCTTC CNNTTCTCTC TCCTCANNCA TATCNANTNT NTTCTACTCT
631 CGTNCNNTAT CTANNCTCCT NTTTCNGTCC TNCTTCTCCT NTCATTTCTA TATTNCTTCT CANACANTNT
701 TCGCATCGTN GCANCATCTC CTCCCATCTC CTGTNCNCTN TTCCN

SEQ. ID NO:22

1 GNNNTTAANT CATTCCCCNC TCNATGCATG CTCGAGCGGC CGCCAGNGTG ATGGATATCT GCAGAATTCG
71 CCCTTGTTTC GGAGGCAGTA GATGAATGGG TTGATGGAAT CTGAGACAGT GCTCTAGAAT CTGTGTTTCA
141 TACAGGATGA GATATAAATG AAACAAATGC TAAATAATGA CACAAGGTAC CTTGCCGAGA GAGGAATCAT
211 CCACCTGGAA GGGTAGGCTG TTTGTGAATA ATGTAGGGTG GGAGAGAAGG CTTTACTAAG GAGATGGGCT
281 TAAAGAATGT GAACGATGTG CTCACAGAGG CCACAGAAGA GAAATTATAG CCAGGAGAAC AACCTGAAAG
351 ACAAAGGACA CGGTGGCATG AGCGCATGTA ACACAATGTA CTCAGGAAAT GGCTGGCATC CTGAGATATG
421 GAGTGGAATA CAGTACAGGG CTTTGTAAAC TCAGCTTGGA GTCAGATCAC AGAAAGCCTT GACAAGGAAC
491 TGAAAATGGG TTCTGAAGGC CAGAAGCCCA TTCAAGATTC CCAAAGGGAA AAACACAAAT CAGCTTGGTT
561 TCAGGACGTA ATTCTTGGCA GTTGCTAGAA TTACATCAGA AAGGAGGTTC ACNT

SEQ. ID NO:23

1 GNNNTNANTC ANNCANTGGG CCCTCTAGAT GCATGCTCGA GCGGCCGCCA GTGTGATGGA TATCTGCAGA
71 ATTCGCCCTT CCTATGTATT TCCTCTTACT GGGCTTTCCT GGTTCTCAAA CTCTTCAGCT CTCTCTCTTT
141 ATGCTTTTTC TGGTGATGTA CATCCCCACA GTTAGTGGTA ATGTGGCTAT CTTGATGTTG GTGAGCACCT
211 CCCATCAGTT GCATACCCCC ATGTACTTCT TTCTGAGCAA CCTCTCCTTC CTGGAGATTT GGTATACCAC
281 AGCAGCAGTG CCCAAAGCAC TGGCCATCCT ACTGGGGAGA AGTCAGACCA TATCATTTAC AAGCTGTCTT
351 TTGCAGATGT ACTTTGTTAT CTCATTAGGC TGCACAGAGT ACTTCCTCCT GGCAGCCATG GCTTATGACC
421 GCTGTCTTGC CATCTGCTAT CCTTTACACT ACGGAGCCAT CATGAGTAGC CTGCTCTCAG CGCAGCTGGC
491 CCTGGGCTCC TGGGTGNGGG GGTTCGTGGC CATTGCAAGT GCCCACAAGC CCTAATCAGT GGCCCTGTCC
561 NTCTGGGGGC CCCCGGGCCA TTNACCACTT TNTTCTGGGA CAATTGCACC CCTGGAATTG G

SEQ. ID NO:24

1 TNNTTAANTC ATTCCNTTGN CCCTCNAGAT GCATGCTCGA GCGGCCGCCA GTGTGATGGA TATCTGCAGA
71 ATTCGCCCTT TCCTTGTTAC TGAGGGAGTA GATTAGGGGA TTGATGGAAT CTGAGACAGT GCTCTAGAAT
141 CTGTGTTTCA TACAGGATGA GATATAAATG AAACAAATGC TAAATAATGA CACAAGGTAC CTTGCCGAGA
211 GAGGAATCAT CCACCTGGAA GGGTAGGCTG TTTGTGAATA ATGTAGGGTG GGAGAGAAGG CTTTACTAAG
281 GAGATGGGCT TAAAGAATGT GAACGATGTG CTCACAGAGG CCACAGAAGA GAAATTATAG CCAGGAGAAC
351 AACCTGAAAG ACAAAGGACA CGGTGGCATA AGCGCATGTA ACACAATGTA CTCAGGAAAT GGCTGGCATC
421 CTGAGATATG GAGTGGAATA CAGTACAGGG CTTTGTAAAC TCAGCTTGGA GTCAGATCAC AGAAAGCCTT
491 GACAAGGAAC TGAAAATGGG TTCTGAAGGC CAGAAGCCAT TCAAGATTCC CAAAGGGAAA AACACANATC
561 ACTTGTTTTC AGGACGTATT CTTGGGCAGT TGCTAGAATT ACATCAGAAA GG

SEQ. ID NO:25

1 GNNNNTTANT CCATGCCCCT CTAGATGCAT GCTCGAGCGG CCGCCAGTGT GATGGATATC TGCAGAATTC
71 GCCCTTGTTT CGCAGCCTAT AAATGAAGGG GTTGATGGAA TCTGAGACAG TGCTCTAGAA TCTGTGTTTC
141 ATACAGGATG AGATATAAAT GAAACAAATG CTAAATAATG ACACAAGGTA CCTTGCCGAG AGAGGAATCA
211 TCCACCTGGA AGGGTAGGCT GTTTGTGAAT AATGTAGGGT GGGAGAGAGG GCTTTACTAA GGAGATGGGC
281 TTAAAGAATG TGAACGATGT GCTCACAGAG GCCACAGAAG AGAAATTATA GCCAGGAGAA CAACCTGAAA
351 GACAAAGGAC ACCGGTGGCA TAAGCACATG TAACACAATG TACTCAGGAA ATGGCTGGCA TCCTGAGGTA
421 TGGAGTGGAA TACAGTACCG GGGCTTTGTA AACTCAGCTT GGAGTCAGAT CCAGAAAGCC CTTGACAAGG
491 AACTGAAAAT TGGGTTCTTG AAGGCCAGAA GCCATTCAAG GATTCCCCAA AGGGGAAAAA CACAAATCAA

561 GCTTGTTTTC AGGGACCGTT AATTCTGGGG CCAGGTTGCT TGAATTACCT TCANGAAAGG GAGGTTCACA
631 CT

SEQ. ID NO:26
1 GNNCTTATTC ATCCCCCTCT AGATGCATGC TCGAGCGGCC GCCAGTGTGA TGGATATCTG CAGAATTCGC
71 CCTTTCTTTG TTCCTCAGAG TGTAGATTAG GGGGTTGATG GGGTTGATGG AATCTGAGAC AGTGCTCTAG
141 AATCTGTGTT TCATACAGGA TGAGATATAA ATGAAACAAA TGCTAAATAA TGACACAAGG TACCTTGCCG
211 AGAGAGGAAT CATCCACCTG GAAGGGTAGG CTGTTTGTGA ATAATGTAGG GTGGGAGAGA AGGCTTTACT
281 AAGGAGATGG GCTTAAAGAA TGTGAACGAT GTGCTCACAG AGGCCACAGA AGAGAAATTA TAGCCAGGAG
351 AACAACCTGA AAGACAAAGG ACACGGTGGC ATAAGCGCAT GTAACACAAT GTACTCAGGA AATGGCTGNC
421 ATNCTNAGAT ATGGAGNGNG AATACCAGTA CANGGCTTTN TANACTCANC TTGGAGTNCA GAATCACANA
491 ANGCCTTGCA AGGAACTGAA AATGGGTTCT GAAAGGCCAG AAGCCNTTNA AGATTCCCAA AGGGAAAAAA
561 CACAAATCAA GCTTTTTTNA AGNACNGTAA TTCNTGGNGC CAGTTGCTTA GAATTNCCAT CANAAANG

SEQ. ID NO:27
1 GGNNTAAGCC TTCCCCCTNC GATGCTGCTC GAGCGGCCGC CAGTGTGATG GATATCTGCA GAATTCGCCC
71 TTCCCATGTA TTTCCTCTTA CTGGGCTTTC CTGGTTCTCA AACTCTTCAG CTCTCTCTCT TTATGCTTTT
141 TCTGGTGATG TACATCCTCA CAGTTAGNGG TAATGGGGCT ATCTTGATGN TGGTGAGCAC CNCCCATCAG
211 TTGCATACCC CCATGTACTT CTTTCTGAGC AACCNNTCCN TCCTGGAGAN TTTGGNATAC CACACGCAAN
281 NAGNGNCCNA AGGCACTTGG NCNTNCTACA GGNGGAGAAG GCTTGACCAT ANNATTTTAC CATGCCTNGC
351 CTTANGNCAN ACCCNNCTTN TNCCTNTTNT TCCNCTNNNN GGTNNNTCAN CCGCANNCTT NNATCNNNTG
421 NANCTTCATN GAATATGGNN TNNGTNTNTC TTGAGAGCCT CNNGATCNNA TTTTTTCCAN CNNCTAAAGN
491 GGNGCTTNTC TCTCTNNNAT CTAGCTTNNT GGNTCTCTTT TNTNTNCTNA CCCGTGNTNT CCTATNTGNT
561 GTCTCTTCCT ACNNNCTGCN NTTATTNTAN ATCANNTCTN NCNTTGCTCT CNTNTACNAC ATNATCATNC
631 TCNCTCCCCN CTNTCNCTCT CTATNNCNTA CCATCNCTCT CTTCTCATTC ANCTCTTTNT CATTGNTTGT
701 TCANTTANNC ACTCTCCNTC NCATCTTCTA TNCACTANNT TTNTTNTTTT NCTCTCTANT TCTNNTTCCA
771 NTGTNCACTC CNNTCTTNNC NNTTNCCCTA NCG

SEQ. ID NO:28
1 GTNNTTNANN NCATTGCCCC TCTNGATGCA TGCTCGAGCG GCCGCCAGTG TGATGGATAT CTGCAGAATT
71 CGCCCTTCCT ATGTACTTCC TCTTACCGGG CTTTCCTGGT TCTCAAACTC TTCAGCTCTC TCTCTTTATG
141 CTTTTTCTGG TGATGTACAT CCTCACGGTT AGTGGTAATG TGGCTATCTT GATGTTGGTG AGCACCTCCC
211 ATCAGTTGCA TACCCCCATG TACTTCTTTC TGAGCAACCT CTCCTTCCTG GAGATTTGGT ATACCACAGC
281 AGCAGTGCCC AAAGCACTGG CCATCCTACT GGGGAGAAGT CAGACCATAT CATTTACAAG CTGTCTTTTG
351 CAGATGTACT TTGTTTTCTC ATTAGGCTGC ACAGAGTACT TCCTCCTGGC AGCCATGGCT TATGACCGCT
421 GTCTTGCCAT CTGCTATCCT TTACACTACG GAGCCATCAT GAGTAGCCTG CTCTCAGCGC AGCTGGCCCT
491 GGGCTTCTGG GTGGGTGGGT TTCGGGGCCA TTGCAAGTGC CCACAGCCCT TATCAAGTGG CCTGTCCTTC
561 TGNGGCCCCC GGGCCCATCA ACCACTTTTT TCTGGGGACA ATTGCACCCT GGAATGGCCC

SEQ. ID NO:29
1 GTNNNTNNNT CCATNCCATT GGGCCCTCTA GATGCATGCT CGAGCGGCCG CCAGTGTGAT GGATATCTGC
71 AGAATTCGCC CTTTCATGGT TCCGGAAACA GTAAATTATG GGGTTCAGTC ATGGTAACAG GAGGAGGCTG
141 AGTGTATGGG CATGGATGGG GGCTGTGAAT GTGGCGGGAG CTCATGGATG TGCTCTTCTG AGTGCTTCAC
211 GTTTCTGAGT GAAATAAGAA GCAAGGTCAT CACCGAGAGG GAGGAGACAG GCTCGGGTGA GTTTAGTGGA
281 TATGAATCCA AGAGAGACCA TTCAACTTAG TTGTCTATTT TTTTTTTCTC CAGTTATAGT CACTTGCATG
351 AATGTAGATG TGGAGTACTT GATCATAAGA TCCATTTTAT GGCAGAAGAC ATTATTTTTC TGAGCCTTCT
421 GCTGTCAGTT TCTAAATAAG CAGGCCAGCC GGGCTGTGCA CCTAAATGTC TGTCTGGGAG GAGCAGGCTG
491 AGAAGTCTTG CAGTCTGCAG GACACCCGAG GAATCGTATT GTGGGAACCG TCCCCGAGAA CCACACGAGC
561 CGTGCTNCTC AGTNCTGACT GGAANAATGA AATTGNAAGC CAAGTNGTTC NNGGANCNNT

SEQ. ID NO:30
1 GNNNTTNANN CCATTGCGCC CTCTAGATGC ATGCTCGAGC GGCCGCCAGT GTGATGGATA TCTGCAGAAT
71 TCGCCCTTCC TATGTATTTC TCTTCCTAAC GATTGGAATG CCTGGGATTA GGCAGATGAT TTTCTTTTTC
141 CCCCATACCC CTCTATTATT TAGGTGATTG AGTTTAAATC CCTTTATCTA CACCCTTCGG AACAAGGGCG
211 AATTCCAGCA CACTGGCGGC CGTTACTAGT GGATCCGAGC TCGGTACCAA GCTTGATGCA TAGCTTGAGT

281 ATTCTAACGC GTCACCTAAA TAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC
351 CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG
421 CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT
491 TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC TTCCGCTTTC TCGCTCACTG
561 ACTCGCTGGG CTTCGGTCGN TCGGCTGCGG CGAGCGGGAT CAGCTCACTC AAAAGG

SEQ. ID NO:31
1 GNNNNNNNNT CANGCCATTG GGCCCTCTAG ATGCATGCTC GAGCGGCCGC CAGTGTGATG GATATCTGCA
71 GAATTCGCCC TTCCTATGTA TTTCTCTTCA CTTTCTCCGA CATCACTCAC AGCCACCCCA CCCTCAGCCT
141 CTCCCTCCTC CCATGTATTT TCTCTTCAAT CTCTCCTTCT TTGATATCCT GAACTTTCTG TAGCTCTTTA
211 TTTTCTCTTC CAATCCCTTC ATATACACGT TTCGTAACAA GGGCGAATTC CAGCACACTG GCGGCCGTTA
281 CTAGTGGATC CGAGCTCGGT ACCAAGCTTG ATGCATAGCT TGAGTATTCT AACGCGTCAC CTAAATAGCT
351 TGGCGTAATC ATGGTCATAG CTGTTTCCTG TGTGAAATTG TTATCCGCTC ACAATTCCAC ACAACATACG
421 AGCCGGAAGC ATAAAGTGTA AAGCCTGGGG TGCCTAATGA GTGAGCTAAC TCACATTAAT TGCGTGCGCT
491 CACTGGCCGC TTTCCANGTC GGGAAACCTG TCGGCCAGCT GCATTAAATG AATCGGCCAA CGCNCCGGGA
561 GAGGCGGTTT GCGTATTGGG CGCTNTTTCG TTCTTCGNTN ACTGATCGNT GG

SEQ. ID NO:32
1 GNNNNNNNNT TCATNCCATT GGGCCCTCTA GATGCATGCT CGAGCGGCCG CCAGTGTGAT GGATATCTGC
71 AGAATTCGCC CTTGTTGCTT AGAGTGTAAA TAAAAGGGTT AACATTGGCT TAGAGGTGAA GAGTAAATAC
141 ATAGGAAGGG CGAATTCCAG CACACTGGCG GCCGTTACTA GTGGATCCGA GCTCGGTACC AAGCTTGATG
211 CATAGCTTGA GTATTCTAAC GCGTCACCTA AATAGCTTGG CGTAATCATG GTCATAGCTG TTTCCTGTGT
281 GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC
351 CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG AAACCTGTCG
421 TGCCAGCTGC ATTAATGAAT CGGCCAACGC GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT
491 CCTCGCTCAC TGACTCGCTG CGCTCGGTCG NTCGGCTGCG GCGAGCGGTA TCAAGCTCAC TCAAAGGCGG
561 TAATACGGTT ATCCACAGAA TCAGGGGGAT ACGCANGAAA GAACATGTGA GCAAAT

SEQ. ID NO:33
1 GNTNTNANTC ATGCCCCCNC CGATGCNTGC NCGAGCGGCC GCCAGTGTGA TGGATATCTG CAGAATTCGC
71 CCTTGTTGCG GAGCGAATAT ATGAAGGGGT TAAGGGAAGA GAAAATACAT AGGAAGGGCG AATTCCAGCA
141 CACTGGCGGC CGTTACTAGT GGATCCGAGC TCGGTACCAA GCTTGATGCA TAGCTTGAGT ATTCTAACGC
211 GTCACCTAAA TAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT
281 TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA
351 TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT TAATGAATCG
421 GCCAACGCGC CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CTTNCGCTTC CTCGCTCACT GACTCGCTTG
491 CGCTCGGTCC GTTCGGCTGC GGCGAGCGGT ATCAANTCAC TCAAAAGGCG GGAATACGGG TTTNCACAGA
561 AATCAGGGGG ATAACGCNGG AAAGAACATG TGAGCCANAN GGCAGCAAAA GGGCNAGGAA T

SEQ. ID NO:34
1 GNNNNNNNNT CANNCCATTG GGCCCTCTAG ATGCATGCTC GAGCGGCCGC CAGTGTGATG GATATCTGCA
71 GAATTCGCCC TTGTTCCGAA GGCTATAGAT GAAGGGGTTT TAGGTTTTTA GGAACACAGG CTAAGGGGGA
141 AGAGAAAATA CATGGGAAGG GCGAATTCCA GCACACTGGC GGCCGTTACT AGTGGATCCG AGCTCGGTAC
211 CAAGCTTGAT GCATAGCTTG AGTATTCTAA CGCGTCACCT AAATAGCTTG GCGTAATCAT GGTCATAGCT
281 GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG CCGGAAGCAT AAAGTGTAAA
351 GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG
421 GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG
491 CTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTCGGCTGCG GCGAGCGGTA TCAGCTCACT
561 CAAAGGCGGT AATACGGGTA TCCACAGAAT CANGGGATAA CGCAGGAAAA GACA

SEQ. ID NO:35
1 GGNNTTNANT CATTGCCCCG CTNGATGCAT GCTCGAGCGG CCGCCAGTGT GATGGATATC TGCAGAATTC
71 GCCCTTCCGA TGTATTTTCT TCTACGTTAA GGTATTTTAA ATTGTTACTA ATGCATAAGG GCAACACATT
141 CTGTAATGCT GACAAGATGA AAGAGCCAAA AGTAATTAAT GATGCTGTTA CCTCACAAAT ATGTATGTGT
211 GGATGTATAT ATATCTATTC AATATATGTA ACTATACATA TGTCTGTTTC TAATTGAAAA CACCAGGTAA

281 TTATCATCTG TAGAAACCCT AGTGTCTCAG ATAAGTTGGC TAGTTTTTTG TTTCACATAA AGGAACAAAC
351 ATTTATAGAT TTATATGTAT ATTAAAAATG GTAAAAATTG GCTGGGTGCA GTGGTTCATG CCTATAATAC
421 CAGCACTTTG GGAAGCCGAG GTGGGCGGAT TACTTGAGGT AAGGAGCCCA GCCTGACCAA CAAGGTGAAA
491 CCCCATCCCT ACTAAAAATA CAAGAATTAG CCCGGGGATG GTGGTGGCCA CCTGTAATCC CAGCTACTTG
561 GGAGACTGAA GCCAGGAAAA TCACTTGACC CAGGAAGCNG AGGTTGCAGG NGAG

SEQ. ID NO:36
1 NGNNNTTGAN TCAATTCNNN GNCGANGCAT GCTCGAGCGG CCGCCAGTGT GATGGATATC TGCAGAATTC
71 GCCCTTCCTA TGTATTTCCT TCTAGCCAAC CTCCCACTCA TTGATCTGTC TCTGTCTTCA GTCATAGCCC
141 CCAAGATGAT TACTGACTTT TTCAGCCAGC GCAAAGTCAT CTCTTTCAAG GGCTGCCTTG TTCAGATATT
211 TCTCCTTCAC TTCTTTGGTG GGAGTGAGAT GGTGATCCTC ATAGCCATGG GCTTTGACAG ATATATAGCA
281 ATATGCAAAC CCCTACACTA CACTACAATT ATGTGTGGCA ACGCATGTGT CGGCATTATG GCTGTCGCAT
351 GGGGAATTGG CTTTCTCCAT TCGGTGAGCC AGTTGGCCTT TGCCGTGCAC TTACCCTTCT GTGGTCCCAA
421 TGAGGTCGAT AGTTTTTATT GTGACCTTCC TAGGGTAACC AAACTTGCCT GTACAGATAC CTACAGGCTA
491 GATATTATGG TCATTGCTAA CAGTGGTGTG CTCACTGTGT GGTCTTTTGT CTTCTAATCA TCTCATACAC
561 TATCATCCTA ATGACCATCC AGCATTGCCC TTTAGATAAG TCGTNCAAAG G

SEQ. ID NO:37
1 GNNNTNANTC CNNNCCNCNN CTAGATGCAT GCTCGAGCGG CCGCCAGTGT GATGGATATC TGCAGAATTC
71 GCCCTTCCCA TGTATTTGCT TCTCAGCAAC TTGTCCTTCT CTGACCTCTG CTTCTCTTCC GTGACCATTC
141 CCAAGTTGTT ACAGAACATG CAGAACCAGG ACCCATCCAT CCCCTATGCG GACTGCCTGA CCCAAATGTA
211 CTTCTTCCTG TTATTTGGAG ACCTGGAGAA CTTCCTCCTT GTGGCCATGG CCTATGACCG CTATGTGGCC
281 ATCTGCTTCC CCCTGCACTA CACCGCCATC ATGAGCCCCA TGCTCTGTCT CGCCCTGGTG GCGCTGTCCT
351 GGGTGCTGAC CACCTTCCAT GCCATGTTAC ACACTTTACT CATGGCCAGG TTGTGTTTTT GTGCAGACAA
421 TGTGATCCCC CACTTTTTCT GNGATATGTC TGCTCTGCTG AAGCAGGCCT TCTCTGACAC TCGAGTTAAT
491 GAATGGGTGA TATTTATCAT GGGAGGGCTC ATTCTTGTCA TCCCATTCCT ACTCATTCTT GGGTCCTATG
561 CAAGAATTGT CTCCTCATCC TCAAGGTCCC TTTTNTAANG GTATCTGCAA GGCCCT

SEQ. ID NO:38
1 NGNNNNTTNA NTCNANGCCN NGNGCCCTCT AGATGCATGC TCGAGCGGCC GCCAGTGTGA TGGATATCTG
71 CAGAATTCGC CCTTCCAATG TATTTACTTC TCAGCCAGCT CTCCCTTATG GACCTGATGT ACATCTCCAC
141 CACCGTCCCC AAGATGGCGT ACAACTTCCT GTCCGGCCAG AAAGGCATCT CCTTCCTGGG ATGTGGTGTG
211 CAAAGCTTCT TCTTCCTGAC CATGGCGTGT TCTGAAGGCT TACTCCTGAC CTCCATGGCC TACGACCGTT
281 ATTTGGCCAT CTGCCACTCT CTCTATTATC CTATCCGCAT GAGTAAAATG ATGTGTGTGA AGATGATTGG
351 AGGCTCTTGG ACACTGGGGT CCATCAACTC CTTGGCACAC ACAGTCTTTG CCCTTCATAT TCCCTACTGC
421 AGGTCTAGGG CTATTGACCA TTTCTTCTGC GATGTCCCAG CCATGTTGCT TCTTGCTGTA CAGATACTTG
491 GGTCTATGAA TATATGGTTT TTGTAAGGAC AAAGCCTCTT TCTTCTTTTN CCTTTCATTG GCATCACTTC
561 TTCTGNGGGC CGAGTCCTAA TTGCTGGCTA TATAATGCAC TCAAAGGAGG GGAGG

SEQ. ID NO:39
1 TAGNNNNNTT ANNTCANNGC CNNTGNNNGC TCAGATGCAT GCTCGAGCGG CCGCCAGTGT GATGGATATC
71 TGCAGAATTC GCCCTTCCAA TGTATTTTCT TCTCAGCAGG AGAGATATTT ATCCTCACTG CCATGTCCTA
141 TGACCGCTAT GTAGCCATCT GCTGTCCCCT GAACTACGAG GCTGCACAGA GTACTTCCTC CTGGCAGCCA
211 TGGCTTATGA CCGCTGTCTT GCCATCTGCT ATCCTTTACA CTACGGAGCC ATCATGAGTA GCCTGCTCTC
281 AGCGCAGCTG GCCCTGGGCT CCTGGGTCTG TGGTTTCGTG GCCATTGCAG TGGCCACAGC CCTCATCAGT
351 GGCCTGTCCT TCTGTGGCCC CCGTGCCATC AACCACTTCT TCTGTGACAT TGCACCCTGG ATTGCCCTGG
421 CCTGCACCAA CACACAGGCA GTAGAGCTTG TGGCCTTTGT GATTGCTGNT GTGGTTATCC TGAGTTCATG
491 CCTCATCACC CTTGTCTCCT ATGTGTACAT CATCAGCACC ATCCTTAGGA TCCCCTCTGC AGTGGCCGGA
561 GCAAAGCCTT CTCCCGTGCT CCTCGCATCT NAACGNGGTG CTCATTTGGT ATGGG

SEQ. ID NO:40
1 CATGCTCGAG CGGNCGCCAG NGNGATGGAT ATCTGCAGAA TTCGCCCTTC CTATGTATTT GCTTCTCAGC
71 AGGAGAGATA TTTATCCTCA CTGCCATGTC CTATGACCGC TATGTAGCCA TCTGCTGTCC CCTGAACTAC

141 GAGGTGATTC ATGTGCCCAT TAGAGCTTGA GAAGCACTGC TTGGAAGCCC CTTCTGCCAT CAATGAGGCT
211 GCACAGAGTA CTTCCTCCTG GCAGCCATGG CTTATGACCG CTGCCTTGCC ATCTGCTATC CTTTACACTA
281 CGGAGCCATC ATGAGTAGCC TGCTCTCAGC GCAGCTGGCC CTGGGCTCCT GGGTCTGTGG TTTCGTGGCC
351 ATTGCAGTGC CCACAGCCCT CATCAGTGGC CTGTCCTTCT GTGGCCCCCG TGCCATCAAC CACTTCTTCT
421 GTGACATTGC ACCCTGGATT GCCCTGGCCT GCACCAACAC ACAGGCAGTA GAAGCTTGNG GCCTTTGTGA
491 ATTGCTGNTG TGGGTATCCC GAGTTCATGC CTCATCACCC TTGNCTTCTA TGTGTACATC ATCAGGCACC
561 ATTCTCAGGA TCCCTTCTGC AAGNGG

SEQ. ID NO:41
1 ATGGNNNNNN NNTTTNNNAA ANTTTTNCCC ANTTTGGGCC GNCCCCCCCT TCTTTAAGGN AATGGGCCCA
71 TTGGGCCCTT CCCGGAAGGC CCGGGGGCNC CCGGCCCCAA AGGTTTGGGT TGGGAAATGG GGGGAATTTA
141 AATTCCTTTG GGCCAAGGNA AAAATTTTCC NGCCCCCCTT TTTTTCCCCT TTTGGTTTTT ANCCGGGGGA
211 ANGGGGGGGT TGATTAATTA ATCGGGAAGN TNGGGGGGAA NTTTTTTAAA AAAAACCTTG GGGGAAGGTT
281 CCAACCCAAC AAGGTTGGTT TTCCANGGGA CCGTTGGGAC CAGGCTTTTN GAATCAAGAA TCCCAAAGGG
351 CATTCTTTTG GATTAAGGAA NGGTGCCGGG ACCGGTGAAA GGGAAAAAAC TGGTGGACCC CATACCAAAA
421 TGAGAACCAC GGTGAGATGC CGAGGAGCAC GTGGAGAAAG GCTTTGCTTC CGGCCACTGG CAGAGGGGAT
491 CCTGAGGATG GTGCTTGATG ATGTACACAT AGGGAGACAA GGGTGATGAG GCATGAACTC AGGATAACCA
561 CAACAGCNAT CACAAAGGCC ACAAAGCTCT ACTGCCTGTG TGTTGGGTGC AGGCCAGGGC AATCCAGGGG
631 TGCAATGTCA CAAGAAAGAA AGTGGTTGAT GGCACGGGNG GGCCACAGAA GGACAGGCCA CTTGATGAAG
701 GGCTTGTGGG CACTGCAATG GCCACGAAAC CACCAGACCC AGGAACCCAN GGCCAAGCTT GCGCCTGAAG
771 AGCAAGGCTA CTCATGAATG GCTTCCGTAG TNGTAAAGGA TAGCAAGATG GCAAAGGCAA GCCGGTCATN
841 AAGCCATGGC TTGCCNG

SEQ. ID NO:42
1 GNNNTTANNN CATTGCGCCC TCTAGATGCA TGCTCGAGCG GCNCGCCAGT GTGATGGATA TCTGCAGAAT
71 TCGCCCTTGT TGCGCAAGGA GTAGATGAAC GGATTCAGGG CAAGGGAGTG CTGAGGAGAT AGACGGGTAT
141 ACACTGGGCA CAAGTCCATG AGTAATCAAG GCCTGTTATT TAAAAAAAAA AAAAAAAAAG CTTGAACAAT
211 ATAGAATCCC ATTACCCAGA GATAGACTGG ATGGTGAATT AAACTTTCTG GTGAATTTCT TTCCAGATAT
281 CTCTCTATGC ATATGTATAC ACAAGCAATT TTTGGAAGAA AAGATACTTT ATAAGGATAA GCCTGAAAAC
351 TGCAACGAAT GCAATGTGGA GAATGAAGGC AAGATGTGGC GAAGAAGGGC ACCACAATCT GGTGGCTGAG
421 AGAGTGCAAC TGTCACTACA GCTAAAAGGA GAGCTGGAGA AGCTGGTGAG GACAGTAAGA GATGAATCTG
491 GTTTAAGACA CGCTGAGTCT CAAATGCCAT GGCTCCCCTA GGTTGCCTCT TCAGATGTAA ATCTTAAGCT
561 CAAAGCAGGT GGATGAGAAA TCACATTTCA TAGTCCCTGC ACAGACGGCT NTNTTGAGCT

SEQ. ID NO:43
1 GNNNNTTAAN TCATTGCCCC GNNNGANGCA TGCTCGAGCG GCCGCCAGTG TGATGGATAT CTGCAGAATT
71 CGCCCTTCCC ATGTATTTGC TTCTCAGCAA CTTGTCCTTC TCTGACCTCT GCTTCTCTTC CGTGACCATT
141 CCCAAGTTGT TACAGAACAT GCAGAACCAG GACCCATCCA TCCCCTATGC GGACTGCCTG ACCCAAATGT
211 ACTTCTTCCT GTTATTTGGA GACCTGGAGA GCTTCCTCCT TGTGGCCATG GCCTATGACC GCTATGTGGC
281 CATCTGCTTC CCCCTGCACT ACACCGCCAT CATGAGCCCC ATGCTCTGTC TCGCCCTGGT GGCGCTGTCC
351 TGGGTGCTGA CCACCTTCCA TGCCATGTTA CACACTTTAC TCATGGCCAG GTTGTGTTTT TGTGCAGACA
421 ATGTGATCCC CCACTTTTTC TGTGATATGT CTGCTCTGCT GAAGCTGGCC TTCTCTGACA CTCGAGTTAA
491 TGAATGGGTG ATATTTATCA TGGGAGGGCT CATTCTTGCA TCCATTCCTA CTCATCCTTG GGTCCTATGC
561 AAGAAATGCT CCTCATCCTC AAGGCCCTTC TNTAAGGGTA TCTGCAAG

SEQ. ID NO:44
1 GNNNNTNANT CNTGCCCTGN CCCNCNGCNC NNGCGCCGCG GCGGATGGAT ATCTGCAGAA TTCGCCCTTG
71 TTACTAAGAG TATAGATGAA CGGATTCAGG GCAAGGGAGT GCTGAGGAGA TAGACGGGTA TACACTGGGC
141 ACAAGTCCAT GAGTAATCAA GGCCTGTTAT TTAAAAAAAA AAAAAAAGCT TGAACAATAT AGAATCCCAT
211 TACCCAGAGA TAGACTGGAT GGTGAATTAA ACTTTCTGGT GAATTTCTTT CCAGATATCT CTCTATGCAT
281 GTGTATACAC AAGCAATTTT TGGAAGAAAA GATACTTTAT AAGGATAAGC CTGAAAACTG CAACGAATGC
351 AATGTGGAGA ATGAAGGCAA GATGTGGCGA AGAAGGGCAC CACAATCTGG TGGCTGAGAG AGTGCAACTG
421 TCACTACAGC TAAAAGGAGA GCTGGAGAAG CTGGTGAGGA CAGTAAGAGA TGAATCTGGN TTAAGACACG
491 CTGAGTCTCA GATGCCATGG CTTCCCTAGG TTGCCTCTTN CAGATGTAAA TCTTAAGCTC AAAGCANGTG
561 GATGAGAAAT ACACATTTNA TAGTCACCTG CACAGACGGT TTTTTGAT

SEQ. ID NO:45
1 CATGCCCCGT CCCNCNAGNT NCNNGCNCCG CGGCCGCNAN GGATATCTGN ANAATTCGCC CTTCCTATGT
71 ATTTACTTCT CCAACTTCTC CTTCCCATCT CTATCATTAG AACCCATTCA TATACACCCT ACGAAACAAG
141 GGCGAATTCC AGCACACTGG CGGCCGTTAC TAGTGGATCC GAGCTCGGTA CCAAGCTTGA TGCATAGCTT
211 GAGTATTCTA ACGCGTCACC TAAATAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT
281 TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT GCCTAATGAG
351 TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
421 GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCTCGCTCA
491 CTGACTCGCT GCGCTCGGTC GTTCGGCTGN GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG NAATACGGTT
561 ATCCACAAGA ATCAGGGGGA TAACGCAAGA AAAGACATGT GA

SEQ. ID NO:46
1 GNNNTNATTN ATTGCATTGG GCCCTCTAGA TGCATGCTCG AGCGGCCGCC AGTGTGATGG ATATCTGCAG
71 AATTCGCCCT TAGTGAGTAG ATGAAAGGGT TCAGCATGGG GGTCACCACA GTGTACATCA TAGCCATGAC
141 AGTGTCCTTT AGAGTAGAAC TATTAGCTGA TGAGCATAAG TAGAGACCAA TAACGGTTCC ATAGAACAGT
211 GACACCACAG ATAGGTGGGA GCCACAAGTA GAGAAGGCCT TGCAGACACC CTTAGAAGAA GGGACCTTGA
281 GGATGGAGGA GACAATTCTT GCATAGGACC CAAGGATGAG TAGGAATGGG ATGACAAGAA TGAGCCCTCC
351 CATGATAAAC ATCACCCATT CATTAACTCG AGTGTCAGAG AAGGCCAGCT TCAGCAGAGC AGACATATCA
421 CAGAAAAGGT GGGGGATCAC ATTGTCTGCA CAAAAACACA ACCTGGCCAT GAGTAAAGTG TGTAACATGG
491 CATGGAAGGT GGTCAGCACC CAGGACAGCG CCACCAGGGC GAGACAGAGC ATGGGGCTCA TGAGGGCGGT
561 GTAGTGCAGG GGGAAGCAGA TGGCCACATA GCGGTCATAG GCCATGGCCA CAAGGAGGAA

SEQ. ID NO:47
1 CNATGGGCCC TCTAGATGCA TGCTCGAGCG GCCGCCAGTG TGATGGATAT CTGCAGAATT CGCCCTTCCA
71 ATGTATTTGC TTCTCAGCAA CTTGTCCTTC TCTGACCTCT GCTTCTCTTC CGTGACCATT CCCAAGTTGT
141 TACAGAACAT GCAGAACCAG GACCCATCCA TCCCCTATGC GGACTGCCTG ACCCAAATGT ACTTCTTCCT
211 GTTATTTGGA GACCTGGAGA GCTTCCTCCT TGTGGCCATG GCCTATGACC GCTATGTGGC CATCTGCTTC
281 CCCCTGCACT ACACCGCCAT CATGAGCCCC ATGCTCTGTC TCGCCCTGGT GGCGCTGTCC TGGGTGCTGA
351 CCACCTTCCA TGCCATGTTA CACACTTTAC TCATGGCCAG GTTGTGTTTT TGTGCAGACA ATGTGATCCC
421 CCACTTTTTC TGTGATTTGT CTGCTCTGCT GAAGCTGGCC TTCCCTGACA CTCGAGTTAA TGAATGGGTG
491 ATATTTATCA TGGGAGGGCT CATTCTTGTC ATCCCATTCC TACTCAATCC TTGGGTCTAT GCAAGAAATT
561 GTCTTCTTCA TNCTCAANGG CCCTTTCTTC TAANGGTATC TTGCAAG

SEQ. ID NO:48
1 ANNNCCNTNG GAGCTCCAAA GCAGTGGTAA CAACGCAGAG TACGCCCCCT ATGTACTTAC TTTTGTTAAG
71 TCCAACCTCC ATCCTCCTTG GCCTTTTGAT TCAATTGATC ACTCCTTCCT CCTCAAAACA CCTTGTTCAC
141 TCATCCTTTC TCAGTCTCCT TTGTGGATTC TTCCTCATTT ATTTGACCTC TTGCTGGTGA ACCCTTTCAT
211 ATACACTCTC CGTAACAAAG AGGGCGTACT TCTGTCGTCT TGAGCGNACT GATGGNACCC AGCTTTTGTT
281 CCCTTTAGTG AGGGNTAATT GCGCGCTTGG CGNAATCATG GNCATAGCTG NTTNCTGNGN GAAANTGNTA
351 TTTCGNTNAC AATTNCACAC AACATACNAG CCGGGAGCAT AAAGGGNNAA GNCCTGGGGN GCCTAATGAG
421 GGAGCTTACT CACAATAATT GGGGTGNGCC CACTGGCCCC TTTTCAGGCG GGAAAACCTN GCGGGGCCAG
491 CTGGAATAAA TGAATCGGGC CACGCGCCGG GGAGGAGGGC GGGTTNNGGA ATTGGGCGCT TTTTCCNTTT
561 CTNGGTTAAT GGACTNGGTN GGCNNNGTCC GTTCGGTTGG GGGGANCGGN NNT

SEQ. ID NO:49
1 AACGCAGAGT ACCGCCCACT ACGTAATCTG TACATGAAAG GGTTTAAAAG AGACTGGGAA GAGAGGAATT
71 GGCAAGATCA AGCAGAGGCA ACTCCTTCTA GTCCTTCTAG TACCGCAAGG GGCAGATAAA TGGAATGGGT
141 AACACCTAGA GGAAAGTATA CTTGCCAAAA GCAAATNCAT AGGGGGGAGT ACATTATCGG GTTGAAAAAA
211 GTATTCCATG CAGATAAAAA CCAAAAGCAA ATACATCGGG GGCGTACTTC TGTCGTCTTT GAGCGTACTG
281 ATGGTACCCA GCTTTTGNTC CTTTAGTGAG GGTTAATTGC GCGCTTGGCG TAATCATGGT CATAGCTGGT
351 TTCTGTGTGA AATTGTTATC CCGCTCACAA TTCACACAAC ATACGAGCCC GGGAGCATAA AGTGTAAAGC
421 CTGGGGTGCC TAATGAGTGG AGCTTACTTA CATTAAATTG CGTTGCGCTC ACTGGCCGCT TTTCCAAGTC
491 GGGAAACCTG TCGTGNCAGC TTCANTAATG AATCGGCCAA CGCCGCGGGG AGAGGCGGGT TGCGTATTGG
561 GCGCTCTTCC GCTTCTTNGT TNACTGACTT CGG

SEQ. ID NO:50

```
  1 GNNNTTTAAC NCCGGNGCTN CNAGCAGTGG AACAACGCAG AGTACGCCCC CGATGTACTT TCTTTTTCAG
 71 TCTCAAGTCT TCCTCTTCTC CAAAGATTTT GTCTTTTCTA CTACCTGAGC TACCAAATCC CTTGTCATCA
141 ATTTCAATAA CTGTATTCTC TTCATCATTT CAACTTCAAA CGTGTCATCT CAGAACAAGC TTCATGTTAC
211 TTCCAATTTT ATCCTTCTTG TTTGCTGATT CCAAGAATTC CAGTCCCATC TAGGCCCGCA ATGCATTGTT
281 CCTGCCACCC TTTTCATATC CTCAATTCCC TTGTATCATC ACTTTCCTTT TATATAGCAC AGATTCCATG
351 ATTCATAACA ATAATTATGT TTTTTTTTGC ATGTGCTCTT AATTTCCTTT CTTGCTCCTA TTATCTTCTA
421 TCATACTTTT CTGGAAACAC TAATTCTGGT GAAATATACT CTTTGTGGAC TTTGCACTTA TGCTCAGTCA
491 GCTGAAGATG ATGGCTAGAC AAATACTCAC AATCATGCTG ACTGGCCCAA TTTATAGTCA TGACCACCGA
561 TTACAAACCC CTTCATTTAT TCTCCGCAAC AGGGGCGTCT TCTGCGCTTG AGCGTCCGGT GGGG
```

SEQ. ID NO:51

```
  1 GCAGTGGTAA CAACGCAGAG TACGCCCGTT ACGGAGGCTG TAAATAAAGG GGTTGAGGAA GTAAAGTACT
 71 TCACAGTACT GGAGCACACA GCATGTGAAT TTCAGCCAAA GGACAAATGC CTCCAAAAAA AGTTAATTCA
141 CAGTGCAGCA GGGCGAGGCA CTTGTCTTAT TCGCTGGTTC TCACATTGAC CCTGAAAGGA CTTTTTTTTG
211 TTAATCCCAT TTTCACAGAT GGGAAAGGGA CTCTGTATGG TTGTCACTTT TATCCAAAGT CTCATAGCCA
281 GTAAGAAGCT GCCCTCAAAG TCCCTACCCT GTCTTCCATT CGACTATTCT GAGGTTCAGA CCCAGAAACC
351 CCATACCTCT GCCTTATATT TTAATGAAAA GTATGTCTCC AGGTTTATGT GGAGAATAAC CAAGACCTCA
421 GAAACATTTA GTGAAAATCA GAGCTAGAAG GAATCTGTTT TTTTGCGAGT TCAGAGAAAC TGACTTGGAT
491 AAGACATCAA AGTTGTCTTG TGCAGCAAAT TCTCCTCCGG CACATAGTAG GCACTCTGAT AAATTCAAAA
561 AGGCTTCTAA GAAGAGGCAG AAGN
```

SEQ. ID NO:52

```
  1 GTGAANCCAN NNTAANNCCN ATTGGAGCTC CAAGCAGTGG TAACAACGCA GAGTACGCCC CCGATGTAGT
 71 TTCTTCTTTC CTTCCTTCCC TCCTTCCTTC CTTCTTTCCT TTCTCTCTCT CTCCCTCTCC CTCTCCCTCT
141 CCCTCTCTCT CTCCTTTTTC TTCTCCTTCC TCCTCCTCCC CCCAATCCGT TCATGACTTC TTCTTCTTCC
211 TCTTCTTCTT CTTTCTTTCT TTCTTCTTTT TCTCTAAGCA GGATCCTGGG CTGTTCAAAC CAGAGAGCTG
281 TAAGTCTTTT CTTTCCCCAT TACTGTTAGA TCCGTTGAAT CGGCTCCAGA AACCAAACAA GTTAACCCTT
351 GCATTTACAC GTTTCGTAAC GGGCGTACTT CTGTCGTCTT GAGCGTACTG ATGGTACCCA GCTTTTGTTC
421 CCTTTAGTGA GGGTTAATTG CGCGCTTGGC GTAATCATGG TCATAGCTGT TTCCTGTGGG AAATTGTTAT
491 CCGCTCACAA TTCCACACAA CATACGAGCC GGGAGCATAA AAGTGTAAAG CCTGGGGTGC CTNATGAGTG
561 AGCTAACTCA CATTAATTGC GTTGCGCTTA CTGNCCGTTT TCAGTCNGGA AAN
```

SEQ. ID NO:53

```
  1 TNANNCNNNT TAANNCCCAT TGGAGCTCCA AAGCAGTGGT AACAACGCAG AGTACGCCCC CGATGTACTT
 71 GCTTCTTCTT CTTTGGAGTG GCTGAATGCT TCCTCCTGGC TACCATGGCA TATGACCGCT ATGTGGCCAT
141 CTGCAGTCCC TTGCACTACC CAGTCATCAT GAACCAAAGG ACTCGTGCCA AACTGGCTGC TGCCTCCTGG
211 TTCCCAGGCT TTCCTGTAGC TACTGTGCAG ACCACATGGC TCTTCAGTTT TCCATTCTGT GGCACCAACA
281 AGGTGAACCA CTTCTTCTGT GACAGCCCAC CTGTGCTGAG GCTGGTCTGT GCAGACACAG CACTGTTTGA
351 GATCTACGCC ATCGTCGGAA CCATTCTGGT GGTCATGATC CCCTGCTTGC TGATCTTGTG TTCCTATACT
421 CGCATTGCTG CTGCCATCCT CAAGATCCCA TCAGCTAAAG GGAAGAATAA AGCCTTTTCT ACATGTTCCT
491 CACACCTCCT TGGTGGCTCT CTTTTCTATA TATCATTAAG CCTCACCTAC TTCCGGCCTA AATCAAATAA
561 TTCACCTGAG GGCACGAAGC TGCTATCATT GCCTACACTG NTATGACTCC A
```

SEQ. ID NO:54

```
  1 GTTNTTCCAT GGACTCCCAA GCAGTGGTAA CAACGCAGAG TACGCCCCCT ATGTACTTAC TTCTTGCTGG
 71 CTTATCATTT ATAGATATCA TTTATTCTTC ATCCATTTCC CACAGATCGA TTTCAGACTT GTTCTTTGGG
141 AATAATTCCA TATCCTTCCC ATCTTGCTTG GCCCAGCTCT TTACAGAGCG CCTTTTTGGT GGGTCAGAGG
211 TCTTTCTTCT GTTGGTGATG GCCTATGACC TTGCATTACT TGGTTATCAT GAGACAATGG GTGTGTGTTT
281 TGCTGCTGGT AGTGTCCTGG GTTGGAGGAT TTCTGCACTC AGTATTTCAA CTTAGTGTTA TTTATGGGCT
351 CCCATTCTGT GACCTCAATG TCATTGATCA TTTTTTCTGT GATATGCACC CTTTATTGAA ACTGGTCTGT
421 ACCGATACCC ATGTTATTGG CCTCTTAGTG GTGGCAATGG AGGACTAGGT TGCACTATTG GGNTTCTGCT
491 CTTACTCATC TCTTATGGNN CATCTGCACT CTCTAAAGAA CCTTAGTCAG AAAGGGAGGT GAAAAGCCCT
561 CTCAACCTGC AGTTCCACAT AACTGGGGGG TGGTTTCTTC TTTGTN
```

SEQ. ID NO:55
1 TTANNCCNNT TNAATNCCNT TGGAGCTCCA AAGCAGTGGT AACAACGCAG AGTACGCCCC CAATGTACTT
71 GCTTCTTCTT TTTTGGGGCT GCTGAGTGCT GCCTCCTGGC CACCATGGCA TATGACCGCT ACGTGGCCAT
141 CTGTGACCCC TTGCACTACC CAGTCATCAT GGGCCACATA TCCTGTGCCC AGCTGGCAAG CTGCCTCTTG
211 GTTCTCAGGG TTTTCAGTGG CCACTGTGCA AACCACATGG ATTTTCAGTT TCCCTTTTTG TGGCCCCAAC
281 AGGGTGAACC ACTTNTTNTG TGACAGCCCT CCTGTTATTG NACTGGTCTG TGCTGACACC TCTGTGTTTT
351 GAACTGGAGG CTCTTGACAG CCACTGCCTA ATTCATTCTC TTTCCTTTCT TGCTGATCCT GGGATCCTAT
421 TTCGCATTCT CTTCACTATC TTTAAGGATG CCGTCAGCTG AGGGGAAACA TNAGCATTCT NCACCTGTTC
491 CGCCCACCTC TTGGGTGGCT CTCTCTTCTA TAGCACTGGC AATCCTTAAC GTATTTTCCG ACCCCAATTC
561 AAGTGCCTTT TTNTGAGAAG CAAAGAAACT GGTTGTCACT TTTTTTTCAC AAGGGGNGAC TTCCAATGTT

SEQ. ID NO:56
1 GNGNTTTNNN CCATGGAGCT CCAAAGCAGT GGTAACAACG CAGAGTACGC CCCCCATGTA CTTTCTTCTT
71 CTTTGGAGTG GCTGAATGCT TCCTCCTGGC TACCATGGCA TATGACCGCT ATGTGGCCAT CTGCAGTCCC
141 TTGCACTACC CAGTCATCAT GAACCAAAGG ACTCGTGCCA AACTGGCTGC TACCTCCTGG TTCCCAGGCT
211 TTCCTGTAGC TACTGTGCAG ACCACATGGC TCTTCAGTTT TCCATTCTGT GGCACCAACA AGGTGAACCA
281 CTTCTTCTGT GACAGCCCAC CTGTGCTGAG GCTGGTCTGT GCAGACACAG CACTCTTTGA GATCTACGCC
351 ATCGTCGGAA CCATTCTGGT GGTCATGATC CCCTGCTTGC TGATCTTGTG TTCCTATACT CACATTGCTG
421 CTGCCATCCT CAAGGTCCCA TCAGCTAAAG GGAAGAATAA AGCCTTTTCT ACATGTTCCT CACACCTCCT
491 TGNTGTCTCT CTTTTCTATA TATCATTAAG CCTCACCTAC TTCCGGCCTA AATCAAATAA TTCACCTGAG
561 GGCAAGAAGC TGCTATCATT GNCCTACACT GTTATGACTC CATGTTGAAC CCCATAATTT ATTCATTCAG
631 C

SEQ. ID NO:57
1 TTATNNCCAT TGGAGCTCCA AAGCAGTGGT AACAACCGCA GAGTACGCCC CCCATGTATT TTCTTTTTCT
71 TGGGGNAGCT GNATGCTTCC TNCTGGCTAC CATGGNATAT GACCGGCTAT GNGGNCATCT GCAGTCCCTT
141 GNNCTCCCAG TCATTATGAA CCAAAGGACA CGGGCCAAAC TGGCTGGTGN TTCCTGGGTC CCAAGCTTTC
211 CTGNAGCTAC TGNGCAAGAC CACAATGGCT CTTNAGNTTT CCATTCTGNG GCACCAACAA GGTGAACCAC
281 TTNTTTCTGN GACAGCCGGC TGTGCTGAAA GCTGGTCTGN TGCAAGACAC AGCACTGTTT GAGATCTACG
351 CCATCGTCGG AACCATTCTG GTGGTCAATG AACCCCTGCT TGCTGATCTT GNGTTCCTAT ACTCGNATTG
421 GTGCTGCTAT CCCTCAAGAA CCCATCAAGC TAAANGGGAA GCAATAAAGN CCTTTCTCTA CGTGCTCCTT
491 AACACCTCCC TTGGTGGCCT CTCTTTTCTA ATATAATCNT CTAAGCCTCA ACCTACTTCT TGGGCCTNAA
561 NTCAAATAAA TTCTTCTGGA GAGGCAAGAA GGTGGTATTC ATTTATNCTA CACTGGTNGN GACTCCATGN
631 TGGAACT

SEQ. ID NO:58
1 GTNATNCCNT TTAATNCCNT TGGAGCTCCA AGCAGTGGTA ACAACGCAGA GTACGCCCGT TCCTCAGACA
71 GTATATGAAT GGGTTAAAAA TGGGCCAGAG CAGATGCAGG AAGATCAAAT AGGAGGCTAC TGCAGTAGAG
141 TCAAATCTAG GGCTGATGGT TTCTTGGGAT GCATAGTAAT AGGTAGATAG AGAAAGTCTT TAGGAGGTAG
211 AATGGACAGG ACTTCACAAT GCATTAAATG TAGGGAGAAA AAAAATGATT CCTGGGTTTC TAGCTTGAGC
281 TAGTAGGGAT AGTGGTAGAA TTTACTGATA TGGAAAACTG GAGGAAAAAG AGTTTGGAAG AGAAAGATGG
351 CAAGTTAAAT ACCTGTGGGA AATATAATCA CAGACACTAA ATAGGCAGCT GTGTGGGTGG CAAAGGAGAG
421 CCATGGGCTA GGAACATACA GTGGGATTCC CTGGCATGTC ATTGGTTACT GAAGTCAGAG TGTATGAGAC
491 AGCCTAAGGA GAGAATNCAC ACAGGAGAAG AAAGAACTAA ACATTCAGTG GCTGGCCAGA GGATGAGAAA
561 CCCAAGAGAT TGGACTGTTT AGGAGCAACA GTGTTGNGAA AAGGGAGAAA NGGTTGAAAT T

SEQ. ID NO:59
1 GGNTTTANNC NCTGGAGCTC CAAAGCAGNG GTAACAACGC AGAGTACGCC CATTGCGTAG CGTGTACATA
71 AAGGGGTTGG AGCTGAAGGA GGAGATAAAG AAGAAGACAG CCAGAACCTT GTCCTCTGTC GGAGATCGCA
141 GGGATCTTGG GCCGTAGATA GGTATAAGCA AAGGGTGCAT AGTAGAAAGT CACTACAGTG AGGTGGGTGC
211 TGCAGGTCGA ATAGGCCTTC TTCCTCCCTT CTGCAGAGTG CATGTGGTAG ACAGCAAGGA GAATCCGGCC
281 ATAGGAACAT GCAATACAAA TGAAGGGAAA CACAAGAAAA ATGGTGGTGC TCAAAAACAC CGTGCACTCA
351 TAGACCCAGG TATCCGTGCA GGCTAGGGTC AACATAGCTG GAACATCACA GAAAAAATGA TTGATGGCTC
421 TGGACTTGCA ATATGGGATA CGGAGTGCAT ATACCGTGTG AGCACAAGAG TTGATGGAGC CTATCATCCA

491 AGATCCTGTT ATCATCAGTG CACACACTCT TTTTCTCATA CGGATGAGAT AGTGGAGAGG AAAGCAAATA
561 GCCACATAAC GATCATAGGC CATTGATGTC AGGAGCAGCG CTTCTGCACC TGCTAAAGTC AGGAAGAAGA
631 T

SEQ. ID NO:60
1 TGTTANTCCN NTTTNCTNCC ATTGGAGCTC CCAAGCAGTG GTAACAACGC AGAGTACGCC CTCCTTGTTT
71 CTGAGAGTGT AGATGAAGGG GTTATAGGAG ATAAAGATCA GGGCAATATG TAGGACAAGG ACACAGACAC
141 TGACAACAAA GTTGATTATC TCATTGACAG TGGTGTCTGT GCAGGCCAGC TTCAGCAGGG GTCTCACATC
211 ACAGAAGAAG TGGGAGATGA CAAAGTCATC ACAAAAGGGC AGGCCAAACA TAGATGTTAC TTGGACAATA
281 GCCATGCCCA GGCCAATCCT CAGTGACCCA GATCCCAGTC AGACACAAGC CCTCTTACCT ATGAATACCG
351 TAAGGGGTTG CAGAAGACCA CATAGCAATC ATATCCCATG GCTATGAGAA GAAAGCAGTT GTTGATGCCA
421 AAAGTCACAT AGAAGAGCTG AGTGACACAG CCTTGCATGA CAATAAGCTA GTGAGGATTC AAGAGGCGAG
491 AAAGCATATG GGAGTAATG GCCACCATGT AGCAGGTCTC AGAGATAGAC AGCAATGCTT AGGAAAAAGT
561 ACATGGGCCG TACTTCTGTC GTCTTGAGCG TACTGATGGT ACCCAGCTTT TGTTCCCTTT

SEQ. ID NO:61
1 GTNANNCCNN TGTAGCTCCN AAGCNGAGCT AACAACNNAG AGAACAACGC AGAGTACGCC CCCGATGTAC
71 TTGTTCCTAC TCTTTGCTGG ATTTGAAAAC TTCCTCCTGT CCGTGATGGC CTATGACCGG TTTGTGGCCA
141 TCTGTCACCC CCTGCACTAC ATGGTCATTA TGAACCCTCA CCTCTGTGGA CTGCTGGTTC TAGCATCCTG
211 GACCATGAGT GCTCTGTATT CCTTGCTACA AATCTTAATG GTAGTACGGC TGTCCTTCTG CACAGCCTTA
281 GAAATCCCCC ACTTTTTCTG TGAACTTAAT CAGGTCATCC AACTTGCTTG TTCTGATAGC TTTCTTAATC
351 ACATGGTGAT ATATTTTACA GTTGCGCTGC TGGGTGGAGG TCCCCTCACT GGGATCCTTT ACTCTTACTC
421 TAAGATAATT TCTTCCATAC ATGCAATCTC ATCAGCTCAG GGGAAGTACA AGGCATTTTC ACCTGTGCAT
491 CTCACCTCTC AGTTGTCTCC TTATTTTATG GTGCAATCCT AGGGGTGTAC CTTAGTCTGC TGCCACCCGC
561 AACTCACACT CAAGTGCAAC AGCCTCAGTG ATGTACACTG GGGCACCCCC AT

SEQ. ID NO:62
1 GNNNNNNNAT TTNATGCCNT TNTTGATTCC CNTTNNNNNN NCAAGCAGNG GTAACAACGC AGAGTACGCC
71 CCCTATGTAT TTCTTCCTAA GATCCAAATA TTAAAATAAA AGACAGTCAT CCCACCACTA ACTAAAGTAG
141 TGTTTCCCAC ACTTCTCTAT TAAGAAGCAT GTGAGATACT TGTTACAAAC ATAACATCCT GGTCCCACCC
211 CAAAGCCACT CAATCAAATA CTCCAGGGAA GGGATCTAGG AATTCGTAGG TTTAACGAGT GCCCCAAAAT
281 GATTATTACC TGTTGGAGAA TCTAGGCAAC AATGAATTAA GGAAAGCTCT CTACCATTTG GTACTGGTAC
351 CAGGTTTGAG GATCACAGGG AAGAGGGTAA GCATATCAGA CTAGCAGAGC TGCCAGAACT CGGGCTTTCA
421 AAAGAGAGGT GCCACCCTCT CCCATGTCCA TGTAAGTAGC AAACAACCCT CTCATGTACA CTCTGAGGAA
491 CAAGGGGGCG TACTTCTGTC GTCTTGAGCG TACTGATGGT ACCCAGCTTT TGTCCCTTTA GTGAGGGTTA
561 ATTGCGCGCT TGGCGTAATC ATGGTCATAG CTGTTTCCTG TGTGAAATTG TTATCCGCTC ACAATTCT

SEQ. ID NO:63
1 TGTAGCTCCA AAGCAGTGGT AACAACGCAG AGTACGCCCT CTTGGTTACG TAAGGGAATA GATGATGGGG
71 TTCAGCATGG GGGTGACTAC AGTGTACATG ACAGTGGCCA CACGGTCCCA CTCTGCTCGC GTCGGGACGT
141 GGCCTGGAAG TAGACTGCAA TGACTGTCCT ATAGAAAGAG GCTCACCACA NCCAGGTGGG AGCCACAGGT
211 GGGNCACAAG TCCCGGAGCC TCCCAGAGGC TTGAGGGCAG CTGGAGCACG GGNAAGCTTG NTATGGNCCC
281 ACAAGGAGGC GAGGATGAGC AGNAAGGGAG TGACCACCAC TTGCNGCGCC CTNGGTGAAG ATGAGCAGCT
351 TGGATGTGGT GGNTGTCAGA GCACGAGAGC CTTTAAGAGA GGCTTGGTGG GTCACAGAAG AAGTGGGNGC
421 ACTTTGTGGG AAAGCACAGA AAGGACAAGC GAGCCATGAG CAGGATATAC AGGAGGGAGT TGTCCGTGGG
491 ACACCAGCCA TGCCATTCCA ACCAGGGCTG CGCACATNGC CGGGGACATT CTCGTGGGAT AAGGGAAGGG
561 GTGCCGGATN GGCACGTATC AGTCATAGGC CTTGGNCGCC AGAAGACAGC TTTNAATTTA CCCCAGG

SEQ. ID NO:64
1 GTTANNCCNT NTANCTNCAA NNGAGGTAAC AACGCAGAGT ACGCCCCCCA TGTATTTGCT TCTTGTCCAA
71 CCTGTCCTTT GTAGAGATCT GCTACACCAC CGTTGTGGTG CCCTTGATGC TTTCCAACAT TTTTGGGGCC
141 CAGAAGCCCA TTCCATTGGC TGGATGTGGG GCCCAAATGT TCCTCTTTCT CACACTTGGT GGTGCTGACT
211 GTTTCCTCTT GGCGATCGTG GCCTATGACC GCTATGTGGC CATCTGCCAC CCTTTGCACT ACCCCTCATC
281 ATGACCTGCA GTCTGTGCGT GCAGATGCTG GGCGGCGCTG TGGGCCTGGC CCTCTTCCTC TCCCTGCAGC
351 TCACCGCCTT AATCTTCACC TTGCCCTTCT GCGGCTACCG CCAGGAAATT AACCACTTCC TCTGCGATGT

421 ACCTCCGTCC TGCGCCTGGC CTGCGCTGCA TCCGTGTTCA CCAGGCTGCC TCTATGTCGT GAGCATCCTC
491 GTGCTGACCG TCCCCTTCTT GCTCATCTGC GTCTCCTACG TGTTCATCAC CTGTGCCATC CTGAGCATCC
561 GTTCTGCTGA GGGCCGGCAC CAGGCCTTTT CAACTGCTCT TCCGG

SEQ. ID NO:65
1 TGTAGCTCCN AAGNNGAGNT ANCAACGCAG AGTACGCCCG CGGAATCTAT AGATGAAAGG GTTTGGNGAG
71 TCAGAAGAAG GAAGTACATG GGAGTCATAA CAGTGTAGGA CAATGATGGC AGCTTCTTGC CCTCAGGTGA
141 ATTATTTGAT TTAGGCCGGA AGTAGGTGAG GCTTAATGAT ATATAGAAAA GAGAGACAAC AAGGAGGTGT
211 GAGGAACATG TAGAAAAGGC TTTATTCTTC CCTTTAGCTG ATGGGATCTT GAGGATGGCA GCAGCAATGT
281 GAGTATAGGA ACACAAGATC AGCAAGCGGG GGATCATGAC CACCAGAATG GTTCCGACGA TGGCGTAGAT
351 CTCAAAGAGT GCTGTGTCTG CACAGACCAG CCTCAGCACA GGTGGGCTGT CACAGAAGAA GTGGTTCACC
421 TTGTTGGTGC CACAGAATGG AAAACTGAAG AGCCATGTGG TCTGCACAGT AGCTACAGGA AAGCCTGGGA
491 ACCAGGAGGT AGCAGCCAGT TTGGCACGAG TCCTTTGGTT CATGATGACT GGGTAAGTGC AAGGGACTGC
561 AGATGGCCAC ATAGCCGGTC ATATGCCATT GGTAGCCCAG GANGAAGCT

SEQ. ID NO:66
1 GTTATNCCTT GTTGCTCCCN AGCAGAGGTA ACAACGCAGA GTACGCCCCT ATTTCTCAGA TATANGATGA
71 AGGGGTTCAG AAAAAGAATG AGCAAAGAAA ATCTGGGCCA GGCGGGCATC AAAAGAAATA GTCTTGTGCT
141 CAACCAGAAA GTCTGCAATC ATTTTAGGGG TAGCAGAAGA GGCAACACAT ACGTCTATAA ATGACAGGTT
211 GGCAAGAAGC AAATACATTG GGGGCGTACT TCTGTCGTCT TGAGCGTACT GATGGTACCC AGCTTTTGTT
281 CCCTTTAGTG AGGGTTAATT GCGCGCTTGG CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA
351 TCCGCTCACA ATTCCACACA ACATACGAGC CGGGAGCATA AAGTGTAAAG CCTGGGGTGC CTAATGAGTG
421 AGCTAACTCA CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCAGTCGGGA AACCTGTCGT GCCAGCTGCA
491 TTAATGAATC GGCCAACGCG CCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT CTCGCTCACT
561 GACTCGCTTG CGCTCGGTCG TTCGGCTTGC GGCGAGCGGT ATCAAGCTCA CTCAAAT

SEQ. ID NO:67
1 GGGTTTTACN CTGTGCNCCC CCAGCAGNGG TAACAACGCA GAGTACGCCC TTGTTGCGAA GAAATAAATG
71 AATGGGTTTA AAATAGACGT GAAGATGGTG TAGAATACAG CAAGGACTTT GTCAACTGAG TAACTGCTGA
141 AGGGCCACAC ATAGATGAAA ATACACGATC CAAAGAATAA AGTGACCACA GTGATGTGAG CAGTCAATGT
211 GGAGTGGGCC TTCACCATGC TTACAGAGGA GCGATTCCTA ACTGTAATAA GTATTACAGT GTAGGANACA
281 ACCAANAGGA GAAAGGAACT CAGAGAAAGA AAGCCACCAT CTGCAACTAT TAGTAGGCTG ACAACATAAG
351 TGTCTATGCA GGCTAACTTN GTNGCTAGAG GAAGGTCACA GAAAAAAACT ATCTACCTTA TTAGGACCAC
421 ANAATGGCAG ATTAACCGTG AATGCCAACT GGCTGGTGGT ATGGATGAAG CCCACAAACC AGGAAATGAG
491 GACGAGCACA ACACATACAC AGNAGCTCAT GATTGANATG TAGTGNGGAG GTTTNCTNTN GCTCATANCC
561 GTNTTGCCA TNGNAACTNG GANCACCATT TTACTTGCAG TGNNGGAGNG AACATGAAAT N

SEQ. ID NO:68
1 GTTANNCCNN TTTAATNCNA TGGAGCTCCA AAGCAGTGGT AACAACGCAG AGTACGCCCC CGATGTACTT
71 GTTCCTACTC TTTGCTGGAT TTGAAAACTT CCTCCTGTCC GTGATGGCCT ATGACCGGTT TGTGGCCATC
141 TGTCACCCCC TGCACTACAT GGTCATTATG AACCCTCACC TCTGTGGACT GCTGGTTCTA GCATCCTGGA
211 CCATGAGTGC TCTGTATTCC TTGCTACAAA TCTTAATGGT AGTACGGCTG TCTTCTGCAC AGCCTTAGAA
281 ATCCCCCACT TTTTCTGNGA ACTTAATCAG GTCATCCAAC TTGCTTGTTC TGATAGCTTT CTTAATCACA
351 TGGTGATATA TTTTACAGTT GCGCTGCTGG GTGGAGGTCC CCTCACTGGG ATCCTTTACT CTTACTCTAA
421 GATAATTTCT TCCATACATG CAATCTCATC AGCTTAGGGG AAGNACAAGG CATTTTCCAC CTGTGCATCT
491 CACCTTTCAG TTGCTCCTTA TTTTATGGNG CAATCTAGGG GTGACCTTAG TTTTGCTGNC ACCCGCAACT
561 CACACTTAAG TGCAACAACC TCAGTGATGT ACACTGGGGT CACCCCATGC C

SEQ. ID NO:69
1 GNGNNNCGAG NTTANNCCTT GGACTCCCAG TAGAGCTACN ANGANTNCGC CNAGCGCGCA NTTNNNCCAG
71 GGTNNTNTTN GTATCACCAA TGAATAGAAA ACAGACACCA CCTTGTCCCT GCCTAGCAAG TAGCTGGAGC
141 TGGGTCGCAA GTACACGAAA AGGGCTGTCC CAAACAGCAG AGTCACCACC ATCAGATGCG AGGCACACGT
211 GTTGCAGGCT TTCCATCGGC CCTCTGCTGA AGGGATCTTC AGGACCGCAG ACACTATGTA ACCATAGGAG
281 ATAAGGAGTT GGAGGAACGA TGTTCCTCCG ACGGTGACCA CCACGAGGAA ATTCACCACT TGACTGAGGA
351 AGGTGTCAGA GCAAGACAGA GCCAGGACTG GTGGGAGGTT GCAGAAGAAG TGGTTGATGA TGTTGGGTCC

421 GCAAAAGTGA AGCCTAAATA TGGAGCTGGC CTGGATCAGG GAGCTCAGGA AGCCACCAAC ATATGCCCCA
491 ACCACCATGC GTGTACAGAG GCCCTGGGTC ATGATAGTGG GGTANAGAAG GGGGCTGGAG ATGGCTTGCA
561 TATCGGTCGT ATGCCATAGC AGTCANGAGG AGGCACTCAA GACAGACCCA TGCCGACNAA GAAAT

SEQ. ID NO:70
1 GNNNNTTTTA CCCCTGNNGC ACANAGCAGT GGTNACAACG CNCGAGTACG CCCCCTATGT ATTTTTTCCT
71 ATTCTGGACA CGCTACTCCT GACCGTGATG GCCTATGACC GGTTTGTGGC TGTCTGCCAC CCTCTGCACT
141 ATATGATCAT CATGAACCCC CACCTCTGTG GCCTCCTGGT TTTTGTCACC TGGCTCATTG GTGTCATGAC
211 ATCCCTCCTC CATATTTCTC TGATGATGCA TCTAATCTTC TGTAAAGATT TTGAAATTCC ACATTTTTTC
281 TGCGAACTGA CGTACATCCT CCAGCTGGCC TGCTCTGATA CCTTCCTGAA CAGCACGTTG ATATACTTTA
351 TGACGGGTGT GCTGGGCGTT TTTCCCCTCC TTGGGATCAT TTTCTCTTAT TCACGAATTG CTTCATCCAT
421 AAGGAAGATG TCCTCATCTG GGGGAAAACA AATAGCACTT TCCACCTGTG GGTCTCACCT CTCCGTCGTT
491 TCTTTATTTT ATGGGACAGG CATTGGGGTC CACTTCACTT CTGCGGTGAC TCACCCTTCC CAGAAAATCT
561 CCGTGGCCTC GGTGATGTCA CTGNGGTCAC CCCCATGTTG ACCCTTTCAT TTACACCCTT AGCAAG

SEQ. ID NO:71
1 GNNNNNNNNN GTTNATNCCN NTTTTAATGC CANTNGAGNT AACAACGCAN GAGTACNCCN NNGNGTACGC
71 CCAGGGTTCA ACCNNTGAAT AGAAAACAGA CACCACCTTG TCCCTGCCTA GCAAGTAGCT GGAGCTGGGT
141 CGCAAGTACA CGAAAAGGGC TGTCCCAAAC AGCAGAGTCA CCACCATCAG ATGCGAGGCA CACGTGTTGC
211 AGGCTTTCCA TCGCCCTCTG CTGAAGGGAT CTTCAGGACC GCAGACACTA TGTAACCATA GGAGATAAGG
281 AGTTGGAGGA ACGATGTTCC TCCGACGGTG ACCACCACGA GGAAATTCAC CACTTGACTG AGGAAGGTGT
351 CAGAGCAAGA CAGAGCCAGG ACTGGTGGGG AGGTTGCAAG AAGAAGTGGT TGATGATTGT TGGGTCCCGC
421 AAAAGTGAAA GCCTAAATAT NGAGCTGGCC TGGATCAGGG GAGCTCAGGA AGCCACAACA TATGCCCCAA
491 CCACCATGCG TGTACAGAGG CCCTGGGTCA TGATAGTGGG GGTNGAGAAG GGGGCCTGGA GATGGCTGCA
561 TATCGGTCGT TGCCATAGCA AGTCAGGAGG AGGCACTTCA GACAGACCCA TGCCNCNAAG AAAAAAAACT
631 GNC

SEQ. ID NO:72
1 GNNNNNNNNN NTTNNNNCNN TNACTCCNGC AGTGGTAACA ANNANTACGC NCAGCGCGCA GTTAACCCTC
71 ACTAANGGTA ANNTNAGCTG GAACACATCA NTACGNTCAN GNNNGCNCNA TGACCGGTTT GTGGNCATNT
141 GTCACCCCCT GCACTACATG GGTCATTATG AACCCTCACC TCTGTGGACT GCTGGTTCTA GCATCCTGGA
211 CCATGAGTGC TCTGTATTCC TTGCTACAAA TCTTAATGGT AGTACGGCTG TCCTTCTGCA CAGCCTTAGA
281 AATCCCCCAC TTTTTCTGTG AACTTAATCA GGCATCCAAC TTGCTTGTTC TGATAGCTTT CTTAATCACA
351 TGGTGATATA TTTTACAGGT TGCGCTGCTG GGTGGAGGTC CCCTGACTGG GATCCTTTAC TCTTACTCTA
421 AAGATAATTT CTTNCATACA TGCAATCTCA TCAGCTCAAG GGGAAGTCAA GGCATTTTTC ACCTGTGCAT
491 CTACCCCTCA GTTGCTNCTT ATTTTATGGN GCAATCCTAG GGGTGACCTT AGTTCTGGTG GCACCCGCAA
561 CTACACTCAA TGCACAAGCT CAGTGATGTA CACTGTGGCA CCCATGCTGA ACCN

SEQ. ID NO:73
1 GTNNNNCNNN TTGATTNCCA TTGGAGCTCC AAAGCAGTGG TAACAACGCA GAGTACGCCC CCTATGTATT
71 TTTTCCTATT CTGGACACGC TACTCCTGAC CGGGATGGCC TATGACCGGG TTGNGGCTGG CTGCCACCCT
141 CTGNANTATA TGATCATCAT GAACCCCCAC CTNTGTGGCC TCCNGGTTTT TGNCACCTGG CTCATTGGTG
211 TNATGACATN CCTCCTCCAT ATTTCTCTGA TGATGCATCT AATCTTCTGT AAAGANTTTG AAANTNCACA
281 TTTTTTTNTG CGAACTGACG TACATNCTCC AGCTGGCCTG CTCTGATACC TTCCTGAACA GCACGTTGAT
351 ATACTTTATG ACGGGTGTGC TGGGCGTTTT TCCCTCCTTG GGATCATTTT CTTCTTATTC ACGAATTGNT
421 TTNATCCATA AGGAAGAATG TCCTCATNTG GGGGAAAACA AATAAGCACT TTTNCACCTG TGGGNCTCAA
491 CCTCTTCCGN CGTTTCTTTA TTTTATGGGG ACAGGCATTT GGGGTCCCAC TTTACTTTTT GNGGNGACTC
561 ACCCCTTCCA GAAAANTTTC CGTGGGCNTC NGGGATGTAC ACTGGNGGCA CCCCCATGTT GAACCCTTTT

FIGURE 2

SEQ. ID NO: 111
ggtcccntcg ngtatncntt naccctctga tgctgctcga gcggccggca gggtgatgga 60
tatctgcaga attcgcccctt ctgttacgca ggaatatata aaggggttac tgaggaataa 120
ataaatgggt tactgaggaa taaataaatg ggttactgag gaacaaatac atagggttga 180
aagaactgta aaatagaaaa aggaccttnt gctgctcctc aggatggcgg nacttagggg 240
ccatgtacat gacgatgnng ctgccnntna agagtcccac tnctcancng cctcagcccg 300
ncttttttnct cacnnnccnt ntttntctnc cctcttnnnc tctttncttc ctattccccc 360
cccttccnct cctcccttttt gcntnaccat tgnccctnat ccctttaatt cnntcnntcn 420
tctcccctct attccttcnn tnttcgnctt cantctctnc ctctttctcc cccnctttct 480
ctcntctnct cttcctctng tcatcctngt tcnttctctt ncctanttcc ctctanccctt 540
ntcttattnc tcctctatnc cctctcatct cacntctcnt cctctcntcn tacttnnctc 600
nnctcttccn ctccgtcntc cnctttctct tcntnacgcc acccctcnnn cntnctctct 660
ntctcntcct cactctctcc tctccctncn cntcactntt ctccncctct acntcctatn 720
ctcncnttct nncttnactt tgtcacgctc tcctcctcct ctctacgcac nttttatctc 780
ttatctcncn catcncccctc nnttctncac nctattnact cttttctcnc atactntatn 840
ctcctntcnn cttanatcnc ctcccttctn tnanccnntc actgcn 886

SEQ ID NO: 112
gctgctcgag cgncgcagcg tcggcagtgt nagggnnatn tgccnnnntn gcnnttagat 60
nanaggnntn agtatggggg tgaccacagt ggtacataac tgaggctgtt gcacttgagt 120
gtgagttgcg ggtggcagca gaactaaggt acacccctag gattgcacca taaaataagg 180
agacaactga gaggtgagat gcacaggtgg aaaatgcctt gtacttcccc tgagctgatg 240
agattgcatg tatggaagaa attatcttag agtaagagta aaggatccca gtcaggggac 300
ctccacccag cagcgcaact gtaaaatata tcaccatgtg attaagaaag ctatcagaac 360
aagcaagttg gatgacctga ttaagttcac agaaaaagtg ggggatttct aaggctgtgc 420
agaaggacag ccgtactacc attaagattt gtagcaagga atacagagca ctnatggtcc 480
aggatgccag aaccagcagt cacagagggt gngggtttca tantgncccct gtagngtcag 540
cnncngacna gatggccnca aaccgntctt nggccctcac gnccctggna ggnngttttc 600
tantccacca cnnntnttct nannc 625

SEQ ID NO: 113
catgcnngag caggctcgag cgccggcagn gtgagggata tctgcagaat tcgcccttcc 60
tatgtacttt ttcctgagcg tatacacaat cccatcatgt actggggaga agncagacca 120
tatcattnac aagctgnctt tngcagatgn actttgnttt ctcattaggc tgcncagagt 180
acttcctcct ggcagccatg gcttatgacc gctgtcttgc catctgctat cctttacact 240
acggagccat catgagtagc ctgctctcag cgcagctggc cctgggctcc tgggtgngtg 300
gtttcgcgcn cantgcagcg cccacagccc tcagnagcgg tcttgctctt ctgngncccc 360
cgtgccatta accactnctt tngcngcant gcnccctgca ttgtcttgtc ctgcccacca 420
nacagcagna nancntgngn cnnttngatc gctgntncgc tctcngntct cactccnttc 480
caccttttncc ntcgcattcc nntntccnnc tcgcnctcct gnccnntcnn tctcctcttc 540
tnaacgcgtc ctccgannng nctnnatgnt cgtctcntnn ntgngcnnng ncagcnnnnn 600
nnnccannnn tnngtgcgcc gctcc 625

SEQ ID NO: 114
gnttaagccc tnnccctctn gangcatgct cgagcggccg ccagtgtgat ggatatctgc 60
agaattcgcc cttgttccgc aaacaataga tgaaaggatt aagtgaagga gtgcccaccg 120
catagaagag accaaagaac ttgcccctcc cttgggcata cggatttttg ggctggaggt 180
agacagcnat gactgagctg tagaagaggg tgaccacagt gagatgggag gagcaggtcc 240
caaaggcctt tctccatgct gtggcagagt taatcctcag cactgcctgg gcagtggctc 300
cataagaggc aaggatgagg ctgagaggca caaccacgaa gatgacactg gacacagcca 360
actggatttc attgnaggag gcatctccac aggagagtnc gnatcagaga tggganccctc 420
acataaaaaa gtcatctatc tgntggtggg gacagaatgn ccatgtggag gntnnatgtn 480

cgtntcnnac ctcttatttt tnttnccccct ttctttcgct cnntcccent tntcccnnct 540
cgccanttcc atncncntct ntcnnttttt ttntntnacc ntnttntcat ntcntctctt 600
tattctcttt ctcttgnctc tcccttctct ctcntnttcc canctctccc g 651

SEQ ID NO: 115
ggntctcggt acaanacttg gccctctaga tgcatgctcg agcggccgcc agtgtgatgg 60
atatctgcag aattcgccct tccaatgtat ttattcctgt tatttggaga cctggagagc 120
ttcctccttg tggccatggc ctatgaccgc tatgtggcca tctgcttccc cctgcactac 180
accgccatca tgagccccat gctctgtctc gccctggtgg cgctgacctg ggtgctgacc 240
accttccatg ccatgttaca cactttactc atggccagnt tgtgcttntg tnccnnacna 300
ttgttgntnc cccactnnnc tntgtnntna gtctnctctn ccntnnactg ctcntcctct 360
tnctccnnga gtcctcnggn nncgtngtcg nttncngcnn tcaattgcan tnccnncntc 420
atcctttctt tanttntcca tntnttcact nattnctctt tatccncnnt ntcnccctcc 480
anctcctncn tagcttactn tttcntgctc tccngngctc ancctttttcn ccatatnttc 540
ttctctcncn tntctctcnc tnnnncccnn nnntcntcgt ntctctgctc cntcttnacg 600
tcntnncnct tatttantnt ctcnccnctn tctcnngctc cancgcngta ccngccctat 660
nnnctcntcc gannntgntc atggcatctn cacattnngc cctactatnn ncgatctatn 720
ttcncgncat ntattncaca tccacntgca ctcctactcn ctctctancc nccgtacatc 780
gcnnctacng ntgncnntcn nccgctcntn cggccncnat nnctccactt tntctnggtc 840
cccctctccg 850

SEQ ID NO: 116
gatgcatgct cgagcgcccg cagtgtgatg gatatctgca gaattcgccc ttccaatgta 60
ctttttcctg aagaacctct ctgttttgga tctgtgctac atctcagtca ctgtgcctaa 120
atccatccgt aactccctga ctcgcagaag ctccatctct tatcttggct gtgtggctca 180
agcctatttt ttctctgcct ttgcatctgc tgagctggcc ttccttactg tcatgtctta 240
tgaccgctat gttgccattt gccaccccct ccaatacaga gccgtgatga catcaggagg 300
gtgctatcag atggcagtca ccacctggct aagctgcttt tcctacgcag ccgtccacac 360
tggcaacatg tttcgggagc acgtttgcag atccaatgtg atccaccagt tcttccgtga 420
catccctcag gtgttggccc tggtttcctg ngaggttttc tttgtagagc tttgacccng 480
ccctgagcct caatgcttgg ntctgggatg ctttattccc atgatgatct ccnatttttcc 540
anatcttctn aangggggctc nagaatccct tnaggaccag antcnagcta aaagcctttn 600
cccnnctgct tccccccacg 620

SEQ ID NO: 117
tggcnctcng atgcatgctc gagcggccgc cagtgtgatg gatatctgca gaattcgccc 60
ttccaatgta tttgttcctg ttatttggag acctggagag cttcctcctt gtggccatgg 120
cctatgaccg ctatgtggcc atctgcttcc ccctgcacta caccgccatc atgagcccca 180
tgctctgtct cgccctggtg gcgctgtcct gggtgctgac caccttccac gccatgttac 240
acactttact catggccagg ttgtgttttt gtgcagacaa tgtgatcccc cacttttttct 300
gtgatatgtc tgctctgctg aagctggcct tctctgacac tcgagttaat gaatgggtga 360
tatttatcat gggagggctc attcttgcat cccattccta ctnatccttg ggtcctatgc 420
aagaattgtc tcctccatcc tcaaggtccc ttcttctaag ggtatctgca aggccttctc 480
tacttgtggc tcccaccctg tctgngggtgt cactggttct atggaaccgt tattggtctc 540
tacttatgct cntcagctaa tagttctact ctaaaggaca ctgcatggct atgatgtaca 600
ctgtggtgac cccatgctg aacccctt 628

SEQ ID NO: 118
gatgatgctc gagcggncgc agtgngatgg atatctgcag aattcgccct tcccatgtat 60
ttgttcctga gcaacctctc cttcctggag atttggtata ccacagcagc agtgcccaaa 120
gcactggcca tcctactggg gagaagacag accatatcat ttacaagctg ccttttgcag 180
atgnacnntg ttttctcant angccntaca gngnncatgt ttncgcnngc cntgacttat 240
gacgcgcntn cnncnntatc nnnntntnct ntnacncnac ttcntcatna tntgnncntn 300

nnttcncctn tggcnnctcn nntcncggnc ttncctntgn ncgtcntcnc ccttnggcct 360
gcatctctnc ntnntcctnn ccnncgnnct ntctttcctt cntaccntnt ttctgtntnn 420
tccctcccct ctctgnntgc nnctncnncn catctnnntg ntctgatcnc tntcttnnnt 480
ccatcnngtn ctnttctctc gtntcttctn cncgccncct gcatcactgn gcattatatn 540
cncngtctca tnnctatctt ccgtncttgt cncttctcct ctatgcncga cgtcntnntn 600
tactatcgtc ntcntcnnat tnngncctgt tccnnngcnc ccgnncntcc anntactctc 660
cangntcntc ctnntcctnt ncnctgtcta attcnnctnt accgcntctn gnctcntcct 720
cgtcnntccc nncttcctcc nctcncgnnn ccnttcagct ntcnanttct antnngnncn 780
cnc 783

SEQ ID NO: 119
nntagatgca tgctcgagcg gcccgccagt gtgatggata tctgcagaat tcgcccttcc 60
tatgtatttc ttcctggcca acctgtcctt cttggagacc tggtacatct ctgngactgt 120
gcccaagtta ctgtttagtt tttggtctgc gaacaacagc atctctttca cactctgtat 180
gatacaactg tacttcttca ttgctcncat gngcacagaa tgcgtgcttc tggccgccat 240
ggcctatgac cgntatgtgg ncatctggcg cccactccac tacccaacca taantgagcc 300
atgggctcct gctccnncct cgctntnnna tanngaaccn acagngtagc gncanctccc 360
tgtncgagaa tctacttcat cntnctgcct tannttntgt gggcccaatg tgcntaanca 420
cttngntctg nggacatttn ctccagnant tnaantctct tnctgcnaca aganactgtt 480
cnttancttg annatnttcn ggnacattnt tcctanggnn ttggnacgag cntntctanc 540
accngcactn cncantaant gctncngttc tantcngtgc cattcntgtg nctnccentt 600
tcatngcntn ncctcccncg aaagcnaant aagtnggngt cttnactttc gccccccacn 660
ncatcncant ggcc 674

SEQ ID NO: 120
ggccctctag atgcatgctc gagcggccgc cagtgtgatg gatatctgca gaattcgccc 60
ttcctatgta tttttcctg ttatttggag acctggagag cctcctcctt gtggccatgg 120
cctatgaccg ctatgtggcc atctgcttcc ccctgcacta caccgccatc atgagcccca 180
tgctctgtct cgccctggtg gcgctgtcct gggtgctgac caccttccat gccatgttac 240
acactttact catggccagg ttgtgttttt gtgcagacaa tgtgatcccc cactttttct 300
gtgatatgtc tgctctgctg aagctggcct tctctgacac tcgagttaat gaatgggtng 360
atatttatca tgngagggct cattcttgtc atcccattcc tactcatcct tgggtcctat 420
gcgagaattg tctcctccnt cctcaaaggc cccttcttct aangggtatc tgcaaggcct 480
tctctacttg gtggctcccc ccctgnctgt ggtgtcactg ttcctattgg aaaccgntat 540
tgggactcta cttatgctca tcangctaat agttttactc ttangggaca ctgncaatgg 600
cctntgaagn taccccctggn gtggaccccc atnntngaac ccc 643

SEQ ID NO: 121
ggccctctag atgcatgctc gagcggccgc cagtgtgatg gatatctgca gaattcgccc 60
ttccaatgta ctttttcctg ttatttggag acctggagag cttcctcctt gtggccatgg 120
cctatgaccg ctatgtggcc atctgcttcc ccctgcacta caccgccatc atgagcccca 180
tgctctgtct cgccctggtg gcgctgtcct gggtgctgac caccttccat gccatgttac 240
acactttact catggccagg ttgtgttttt gtgcagacaa tgtgatcccc cactttttct 300
gtgatatgtc tgctctgctg aagctggcct tctctgacac tcgagttaat gaatgggtga 360
tatttatcat gggagggctc attcttgcat cccattccta ctcatccttg ggtcctatgc 420
aagaattgnc tccttccatc tcaaggnccc ttnttctaaa gggtatctgc aaggccttct 480
ctanttgtgg ctcccacccct gtcttgtggn tggcactgnt tctaatggga accggtaatt 540
gnancnctna cnttatgctc natcaactta aatagtttct nactttnaaa gggaccactn 600
ntcattggct tanggatngn ncnttggttt cntggaaatc ccnatcattc ttacnng 657

SEQ ID NO: 122

atgaccctna gatgcatgct cgagcggccg ccagtgtgat ggatatctgc agaattcgcc 60
cttccaatgt atttgttcct gtccaacctg tcctttttgg atattggctt tatctctaca 120
ataattccca atatgctaga tcatattagc tcaggaatta agctgatttc ttatggggag 180
tgtctgacac aactctattt ctctggccta tttgcagatc tggacaacaa ctttctcctg 240
gctgtgttgg cccttgaccg ctatgtggcc atcagccatc ctctccatta tgccctaacc 300
atgaactccc aacgctgtgt cctgttggtg gctgtgtcat gggtgatcac tattttacat 360
gccctagtgc ataccctcct agtgaccagg ctttccttct gtggtccaaa tattatccct 420
cacttcttct gtgatctggc ccactcctg aagctggcct gctccagtac ttgtgtcaat 480
gatctggtgc tcatccttgt ggcaggaaca ctgctgaatg cgcccctttgc tgcatnctta 540
tgncctactt ttacattgca ttggccatcc tgagaattga ttccccnagg ggtatgcaaa 600
gggcccttnt ccagcntcnc nn 622

SEQ ID NO: 123
gcgncgcagt gtgatggata tctgcagaat tcgcccttcc aatgtatttg tttctgttat 60
ttggagacct ggagagcttc ctccttgtgg ccatggccta tgaccgctat gtggccatct 120
gcttccccct gcactacacc gccatcatga gccccatgct ctgtctcgcc ctggtggcgc 180
tgtcctgggt gctgaccacc ttccatgcca tgttacacac tttactcatg gccaggttgt 240
gtttttgtgc agacaatgtg atccccccact ttttctgtga tatgtctgct ctgctgaagc 300
tggccttctc tgacactcga gttaatgaat gggtgatatt tatcatggga gggctcattc 360
ttgtcatccc attcctactc atccttgggt cctatgcaag aattgtctcc tccatcctca 420
aggtcccttc ttctaagggt atctngcaag gccttctcta cttgcggctc cacctgcctg 480
tggtgtcact gttctatgga accgttattg gtctctactt atgctcatca gccaataagt 540
tttactctaa aaggacactt gtcatggnnt atgatgtacn ctgtggngac ccccatgctn 600
aaccccnttn 610

SEQ ID NO: 124
ccttgggccc tctagatgca tgctcgagcg gccgccagtg tgatggatat ctgcagaatt 60
cgccccttctt tattcctgag tgaatatatg agggggttgg cactgctgtt aagagtggac 120
aggaaaatgg aaactagacg aacgtgacaa atccacgtgg atccagaaaa ataggaatca 180
ctgaatgcca aagggcaggt cacagaggag gaagaccagc actctgagca ggatggtcat 240
gtacagcctg gtcaagggca tcttccggga tccacaaagg atcctgacca gcagaaccgg 300
gctggacccg cagagaacca cacataaaaa aatcagccat gtgactgtga tgaaatctga 360
tgtttcacac caaacagaat caagcaccac tagacaggaa gccacagaac atccattcca 420
ggatgctctg cagcagggac agggcccaga gcaggacaca cgactgctna ccaggtnttt 480
tngngtggct gcnagctctn cttaggatng tccccaagga ttgnccnggn ccggtncttt 540
gnttgcttnt cgnnncccta nctatgcctt ngctcctgtn nangcttgac nattggncct 600
cncccacgng gcttaannnt ctcnngncgc atttanancg tnatnntact tcccttgtcg 660

SEQ ID NO: 125
gnccctctag atgcatgctc gagcggccgc cagtgtgatg gatatctgca gaattcgccc 60
ttcctatgta cttcttcctg ttatttggag acctggagag cttcctcctt gtggccatgg 120
cctatgaccg ctatgtggcc atctgcttcc ccctgcacta caccgccatc atgagcccca 180
tgctctgtct cgccctggtg gcgctgtcct gggtgctgac caccttccat gccatgttac 240
acactttact catggccagg ttgtgttttt gtgcagacaa tgtgatcccc cactttttct 300
gtgatatgtc tgctctgctg aagctggcct tctctgacac tcgagttaat gaatgggtga 360
tatttatcat gggagggctc attcttgtca tcccattcct actcatcctt gggtcctatg 420
caagaattgt ctcctccatc ctcaaggtcc cttcttctaa gggtatctgc aaggccttct 480
ctacttgngg ctcccacctg tcttnggngg cactgttcta tgggaaccgg tattggtctc 540
tacttaatgc tcatcaagct aatagttcta ctctaaagga cactgncatg gctatgatgt 600
acactgtggt gaccccccnat gctgacccat tc 632

SEQ ID NO: 126
tctagatgca tgctcgagcg gccgcagtgt gatggatatc tgcagaattc gcccttccaa 60
tgtacttgtt cctggcagcc atggcttatg accgctgtct tgccatctgc tatcctttac 120
actacggagc catcatgagt agcctgctct cagcgcagct ggccctgggc tcctgggtgt 180
gtggtttcgt ggccattgca gtgcccacag ccctcatcag tggcctgtcc ttctgtggcc 240
cccgtgccat caaccacttc ttctgtgaca ttgcaccctg gattgccctg gcctgcacca 300
acacacaggc agtagagctt gtggcctttg ngattgctgg tgtggttatc ctgagttcat 360
gcctcatcac ctttgtctcc tatgtggaca tcatcagcac catccttcag gatcccccttt 420
gncagtgccc ggagnaaaag ncttttccac gtgctcctcg cntctcnncg nggtgctcna 480
tttggtatgg gtccacaagn tnttctttca cgnccggatt ntccattcaa aagatgncct 540
tgnnntttna ncaaaagctt ggncnncgnc ctgaaanact gnngtngact tcangnttta 600
aaactccttt natntcactn ttangggaac naggggcgnn ac 642

SEQ ID NO: 127
ntgngccctc tagatgcatg ctcgagcggc cgccagtgtg atggatatct gcangaattc 60
gcccttccca tgtatttatt ccttagcctg ttggattccc agctgcacag ctggattgtg 120
ttacacaact caccttcttc aagaatgtgg aaanctataa tttttttct gtgacccatc 180
tcaacttctc aaccttgcct gttctgacag catcatcaat aacatattat gtattttaga 240
tatccctata tttggttttc ttcccattnc agggatcctt ttgncttacc atanaattgt 300
cctcctccat tccaagaatt ccattgncag acgggacgna tnangccttc tctacctgtn 360
cntctnaccc gnnagtcgnt tntttatctn tgnantnccc tngggcgncn nccctgncct 420
cagcnttngt cancnttctc cncacnnntt cgtcgntgtt ncccagtnct gtnctnctnc 480
tctcntncnc tttctgcctc ccctccanng tctnnctttc tcagcncccct tnngncncnt 540
gccagcnccn nangntccnc ccctctccct cntgtctnct cnctcctntt cttcntntcc 600
tnnctcatnn nnncgcncnc ncgctctccn cccntntctn tacgactccn gncgtctctn 660
cgcctacgac ctccctgtnc ncnnccgg 688

SEQ ID NO: 128
gcgtgctgcn agcggggcgg cagagtgagc ggatatctgc agaatncgcc cttccgatgn 60
atttctttct aagcaactta tctttcattg acatctgcta ctcttctgct gtggctccca 120
atatgctcac tgacttcttc tgggagcaga agaccatatc atttgtgggc tgtgctgctc 180
agttttttt ctttgtcggc atgggtctgt ctgagtgcct cctcctgact gctatggcat 240
acgaccgata tgcagccatc tccagccccc ttctctaccc cactatcatg acccagggcc 300
tctgtacacg catggtggtt gnggcatatg ttggtggctt cctgagctcc ctgatccagg 360
ccagnnccat atttaggctt cacttttgcg gacccaacat catcaaccac ttcttctgcg 420
acctccacca gtcctggctc tgtcttgctc tgacaccttc cttnagtcaa gncgncgaat 480
tntcccgtgg tgntcacntg tcgngaggaa acatcgnttt cctccaaccc cttantctcc 540
canggg ntac catagngtct gcgngtccct gaagaatcct tttngccaan cgggcgaatn 600
gnaagccctn ccaccgccc 619

SEQ ID NO: 129
gcggcgcagt gtgatgntat ctgacgaatt cgcccttccg atgtatttat ttctaagcaa 60
cttatctttc attgacatct gctactcttc tgctgtggct cccaatatgc tcactgactt 120
cttctgggag cagaagacca tatcatttgt gggctgtgct gctcagtttt ttttctttgt 180
cggcatgggt ctgtctgagt gcctcctcct gactgctatg gcntacgacc gatatgcngc 240
catctccagc cccttctctcn accccactat catgacccag ggcctctgta cacgcatgga 300
ggtngcgccn tatgntngtt gnctncntng agctccctga nccannnctn ntcacntatt 360
ntaggctcna ccnntcgngc tcccgntcca ncancnaacc ccnttcgttc ctgnanactt 420
ctccancacg ttcctggctt ttctgcnntc gcctcncgnc nncc ttatnc ttnangntca 480
cncctganct gcnnnttctt ccangccngc ncgcncancc cgntctntct gnngaancct 540
ttnccatnct gctcnatnct nctctcatcn nttcctantn ctctccnnct cncgctcnnt 600

nncttncnct ctnaacctnt cnnatcctca cctnngatat cctcncgntc tttcgncntc 660
nttcnctgtc cganntcctc anacncntcc ctanncg 697

SEQ ID NO: 130
ctctagatgc atgctcgagc ggccgccagt gtgatggata tctgcagaat tcgcccttcc 60
tatgtattta ttccttagcc acttggccct cactgacatc tccttttcat ctgtcactgt 120
ccctaagatg ctgatgaaca tgcagactca gcacctagcc gtcttttaca agggatgcat 180
ttcacagaca tatttttca tatttttgc tgacttagac agtttcctta tcacttcaat 240
ggcatataac aggtatgtgg ccatctgaca tcctctacat tatgccacca tcatgactca 300
gagccagtgt gtcatgctgg tggctgggtc ctgggtcatc gcttgtgcgt gtgctctttt 360
gcgtaccctc ctcctggccc agctttcctt ctgtgctgac cacatcatcc ctcactactt 420
ctgtgacctt ggtgccctgc tcaagttggc ctgctcagac acctccctca atnagttagc 480
aatctttaca ggagcattga cnggcattat gcttccattc ctgngcatcc tgggttctta 540
tgggcanatn tgggggtcac cattctncag anttccttta ccagggcatn tgcaangcct 600
tggccacttg tggnncccnc tcncg 625

SEQ ID NO: 131
ttggcctcta gatgcatgct cgagcgccgc cagtgtgatg gatatctgca gaattcgccc 60
ttgatacatg attgggttgc ggaaggaata aatcatcggg ttgcggaagg aataaataca 120
tcgggttgcg gaaggaataa atacatcggg ttgcggaagg aataaataca tcgggttgcg 180
gaaggaataa atcatcgggt tgcggaagga ataaatacat cgggttgcgg aaggaataaa 240
tacatcgggt tgcgtaagga ataaatcatt gggttgcgta aggaataaat cattgggttg 300
cgtaaggaat aaatcattgg gttgcgtaag gaataaatca ttgngttgcg taaggaataa 360
atctttgtgc tggtaccgat ctatcatggg gttacgaaag ggaagaaata cattggaang 420
ggcgaattcc agcacactgc cgnccgctac tagtgggatc cganctcggt accaagcttt 480
gatgcntagc ttgagtattt taacgccgcc aacctaaaat ngcnttggcc ttacncnntg 540
gaccnagctt gncttccttg cgtnaanttt cnttattcct cctntntntc ttctcccccc 600
ncanaatnnt nncccngntn ancacncann ttntatannc ctngngctcc cctantc 657

SEQ ID NO: 132
tggcccncta gatgcatgct cgagcgncgc cagngtgang gatatctgca gaattcgccc 60
ttcctatgta tttattcctt aatgtcctct cgcttcttga tatttgttac tcttctgtgg 120
tcacacctaa gctcttggtc aacttcctgg tctctgacaa gtccatctct tttgagggct 180
gtgtggtcca gctcgccttc tttgtagtgc atgtgacagc tgagagcttc ctgctggcct 240
ccatggccta tgaccgcttc ctatccatct gtcaacccct ccattatggt tctatcatga 300
ccagggggac ctgtctccag ctggtagctg tgtcctatgc atttggtgga gccaactccg 360
ctatccagac tggaaatgtc tttgccctgc ctttctgtgg gcccaaccag ctaacacact 420
actactgtga cataccaccc cttctccacc tggcttgtgc caacacagcc acagcaagag 480
nggncctcna tgncttttct gntctggcac ccttctggcn gctgcaggca ttctcacctc 540
taccggcttg ggcttggggg ccaatnggga ggatgcgcct caagaacagg gagggagaaa 600
ggactcccca cttntgcctc ccnn 624

SEQ ID NO: 133
ggagttgata tgaacgggtt aagtgaagga gtgcccactg catagaagag accaaagaac 60
ttgcccctcc cttgggcata cggatttttg ggctggaggt agacagcaat gactgagctg 120
cagaagaggg tgaccacagt gagatgggag gagcaggtcc naaaggcctt tctccatgct 180
gtggnagagn taattctcag cactgcctgg gcagtcggct ncataagagg caaggatgag 240
gctgagaggc acaaccacga agatgacact ggacacangc caactgtatc cattgtagga 300
ggnatctcca caggagagtn gaatcagaga tgggacnttc acattaanaa gttatttatn 360
tgctggcggg nacagatgcc caagcgnnan ggngntatgg tnctggncna ttnnttcgtc 420
canacccatt atctcangcc acatgtatnt cagcntttna ntcncnntnt nagtntagtc 480
tngntgntnt ncnnnattnn ccnntctttn tcccntcann tatcattntc attccttncn 540
ncncanantt atggnncncc cgnacncnct cngtnactcc cctnnngncg 590

SEQ ID NO: 134
gnntnnnnnn ntgttancct cgtccctcta gatgcatgct cgagcggccg ccagtgtgat 60
ggatatctgc agaattcgcc cttccgatgt atttatttct acacagacac agtgacaatc 120
tgatctctct tgcttttccc cacacactgc aacctctgcc tccacattca agtgattctc 180
ctgcctcagc ctcttgagta gctggaatta cagatgtgag ccaccatgcc tggcctgtcc 240
agatgttttt gaaacaaccc ccaccagcac tggagggagt caagggaaga caagccaggc 300
atctgagctc ctctgtctct gcctttcctt ctcactgtcc ccagggtaac ccgtcaccac 360
ccccatcacg aacccccttca tctacacatt acgtaacaag ggcgaattcc agcacactgg 420
cggccgttac tagtggatcc gagctcggta ccaagcttga tgcatagctt gagtattcta 480
acgcntcacc taaatagctt ggcgtnatca tngncccnag cttgntttct gtgtgaaatt 540
tgntatccgc tcacaaattc cacacaacat acgagccnga agcaataagn nntaaagcct 600
gnggtgccna angagngagc taactcacaa ttaattncgt tggctnactt gcccc 655

SEQ ID NO: 135
ttgngccctc tagatgcatg ctcgagcggc cgccagtgtg atggatatct gcagaattcg 60
cccttcctat gtacttgttt ctaagcaacc tctccttcct ggagatttgg tataccacag 120
cagcagtgcc caaagcaccg gccatcctac tggggagaag tcagaccata tcatttacaa 180
gctgtctttt gcagatgtac tttgttttct cattaggctg cacagagtac ttcctcctgg 240
cagccatggc ttatgaccgc tgtcttgcca tctgctatcc tttacactac ggagccatca 300
tgagtagcct gctctcagcg cagctggccc tgggctcctg ggtggtgtgg tttcgtggcc 360
attgcagtgc ccacagccct catcagtggc ctgtccttct gtggttcccg tgccatcaaa 420
cacttcttct gtgacattgc accctggant gccctggcct gcaccaacac cacaggcagn 480
aagagcttgt ggcctttgng aatcgcctgn tggggctanc cttnngtcat gccctnatca 540
ccntttntcn nctatgnngt acantcatta agcnccaatc nctcatggga tccccttttg 600
cnagtggccc ggcgngcnaa ngncctnctc cccgtnccn 639

SEQ ID NO: 136
tgnccctcta gatgcatgct cgagcggccg ccagtgtgat ggatatctgc agaattcgcc 60
cttccgatgt atttgtttct agccaacctg tcattaactg atgcttgntt cacttctgcc 120
tccatcccca aaatgctggc caacattcat acccagagtc agatcatctc gtattctggg 180
tgtctngcac agctatattt cctccttatg tttggnggcc ntgacaactg cctgctggct 240
gtgatgccat angaccgtta tgtggccatt tgccaaccac cccattacag cacatctatg 300
agtccccagc tctgtgcact antgctgcnc gtgtgctgng tgcnanccan ttgtctgcct 360
gctgcacatn ctgttnccnc ccccncgngg nctctttnnn ccgnacccnc cctacaantc 420
cntatcannt tcngctnccc tttcttctcc ccccnnttct tnccnccttc ctcnnnccta 480
ctttcttctc tcnccntnct canatnatca gtccnacctc nccttcnttt cttcactnan 540
tnctctcnct cccnctcacc ngntngtcta gtctgccgtc gccccntcgc tatcnctncc 600
cccctctccg cntcccctga tcgtcctngt ctaccctcnc catctnatcc ctcc ·654

SEQ ID NO: 137
ctctagatgc atgctcgagc ggccgccagt gtgatggata tctgcagaat tcgcccttcc 60
aatgtatttt tttctaagca acctctcctt cctggagatt tggtatacca cagcagcagt 120
gcccaaagca ctggccatcc cactggggag aagtcagacc atatcattta caagctgtct 180
tttgcagatg tactttgttt tctcattagg ctgcacagag tacttcctcc tggcagccat 240
ggcttatgac cgctgtcttg ccatctgcta tcctttacac tacggagcca tcatgagtag 300
cctgctctca gcgcagctgg ccctgggctc ctggncgtgn ggcttngtgn cnttgcngcn 360
ctcctagcnc tcatgnnnnc cttgccttnt gggnccntgn nnatcaccct nttnctctgt 420
nacacttgta cctcncgnct tgccctnnnc tgcttctaan tccctnngtt gtantncctn 480
gccttntctc cccttcgctn gttnatcttn anntncntgc ntctntgncc ctctccttcg 540
ttngacccct ntanncncnc tcttcttcnn anntcccctc tatcncnccg ntnncctccn 600
ntgtcnnccg antangntac ntntcacnnt ntntcnnctn ctctcctaac tcttnccg 658

SEQ ID NO: 138
ggccccctag atgcatgctc gagcgggcgc cagcgtgatg gatatctgca gaattcgccc 60
ttcccatgta tttgtttcta agcaacctct ccttcctgga gatttggtat accacagcag 120
cagtgcccaa agcactggcc atcctactgg ggagaagtca gaccatatca tttacaagct 180
gtcttttgca gatgtacttt gttttctcat taggctgcac agagtacttc ctcctggcag 240
ccatggctta tgaccgctgt cttgccatct gctatccttt acactacgga gccatcatga 300
gtagcctgct ctcagcgcag ctggccctgg gctcctgggt gngtggnttc gtggccantg 360
tagtgcccac agcccntatc agnggcctgt ccttttgtgg ncncccgtnc catcaacccc 420
ttctttctgt gacatttgcc cccctgcntt ncccntggcc ctnccccaan cacngcangg 480
nngnttncnn gnctcggcnc cccctttgac ntantncntt gntgngcgct tatncntgcg 540
tttaatgncc ttaatnaaac tctcnctctt catgttnttc ntttntntng gnaccaantc 600
ttcnaannna ccctttttc catnnncncg tctacntcnc tctcnccttc ntcgngtttn 660
nnngtcnncc 670

SEQ ID NO: 139
gatgcatgct cgagcggccg ccagtgtgat ggatatctgc agaattcgcc cttccgatgt 60
atttttttct aagcaacctc tccttcctgg agatttggta taccacagca gcagtgccca 120
aagcactggc catcctactg gggagaagtc agaccatatc atttacaagc tgtcttttgc 180
agatgtactt tgttttctca ttaggctgca cagagtactt cctcttggca gccatggctt 240
atgaccgctg cttgccatct gctatccttt acactacgga gccatcatga gtagcctgct 300
ctnagcgcag ctgncctggg ctcctgggtg ngtggttcng ngccattcag cgcccacagn 360
cttcatcagt ggncttgtnc ttctgngccc ccgcncatcn aaccantttc ttctgngana 420
atngtacccc tgnanttgcc ctggccttgt anccancaca tangctcgta tgngcttctn 480
ntggccnccn tgnttcgcnt ngtnnccgng ntanccngnc tnnacgtcct ttcnnacact 540
ttnnctctat gttntcaacn tcncngncta ttcgctcang atanccactc ttncanncnt 600
cggannnnta nnctttccnn accntctttc cntnc 635

SEQ ID NO: 140
atgaccctct agatgcatgc tcgagcggcc gccagtgtga tggatatctg cagaattcgc 60
ccttcctatg tatttatttc taagcaacct ctccttcctg gagatttggg tataccacag 120
cagcagtgcc caaagcactg ggccatccta ctggggagaa gtcagaccat atcatttaca 180
agctgtcttt tgcagatgta ctttgttttc tcattaggct gcacagagta cttcctcctg 240
gcagccatgg cttatgaccg ctgtcttgcc atctgctatc ctttacacta cggagccatc 300
atgagtagcc tgctctcagc gcaagctggc ctgggctcct gggtgtgtgg tttcggnggc 360
cattgcagng cccacagcnc tnatcagtgg gctgtccttt ctgtgggccc ccgngcccat 420
tcaacccacn tttctttttg nggatattgg caacccentg gnatttgncc cctnggccct 480
ngcacncaaa ccancaccag ggtcngnnna caanctttgn cgggccccttt ttntgaaatt 540
ggcctnggtg ngggnntaat tcnctttggn tttnaatgcc cttccaatna accttttttgn 600
cnttcntatg ggngnnccct tnnattcnag caccacancc ttangggaac ccncctttt 660
gtcaagtgng nccggtnann naaaagccnt nttccnnntg cccccccg 709

SEQ ID NO: 141
ntgggccctg agatgcangc tcgagcggcc gccagtgtga tggatatctg cagaattcgc 60
ccttcccatg tattttttc taagcaacct ctccttcctg gagatttggt ataccacagc 120
agcagtgccc aaagcactgg ccatcctact ggggagaagt cagaccatat catttacaag 180
ctgtcttttg cagatgtact ttgttttctc attaggctgc acagagtact tcctcctggc 240
agccatggct tatgatcgct gtcttgccat ctgctatcct ttacactacg gagccatcat 300
gagtagcctg ctctcagcgc agctggccct gggctcctgg gtctgtggtt tcgtggccat 360
tgaagtgncc acanngcctc atcagntggc cntgtccttc tgcnnccccc cgtnncattn 420
nncacttctt tcgtgacatt gccannctnn tnttgccctn gtccttnncc natcatccat 480

ggcngttngn gctgttggcc ctttcgctca cncngtctgc gcccattctc nctgtnncaa 540
nngcctccnt ctactctctg cnttctanct antnncncct ctttnctncc tnnantctnt 600
cctcgatctc ctttcangnc tccgctncac tgctcnctna acgtccnttt cttccctnnt 660
nntcnnntnc g 671

SEQ ID NO: 142
gggcnncttt gggtatgcct tgnccccttag atgcatgctc gagcggccgc cagtgtgatg 60
gatatctgca gaattcgccc ttccaatgta cttatttcta gccaacctgt cattaactga 120
tgcctgtttc acttctgcct ccatccccaa aatgctggcc aacattcata cccagagtca 180
gatcatctcg tattctgggt gtcttgcaca gctatatttc ctccttatgt tnggnggcct 240
tgacaactgc ctgctggctg tgatggcata tgaccgctat gtggccatct gccaaccact 300
ccattacagc acatctatga gtccccagct ctgtgcacta atgctgtgng tgngctgngt 360
gctaaccaac tggnctgccc tgatgcacac actgttnctn atcccngcgc tttcttggtc 420
ccnntangcc nctncnntcn ttccnttntn tntctctacc tctcccntcg ngctctnccc 480
cttccccnnt cttcctnntg tactcnctan nnctgttnnn cccccntcntt ctcttcttcc 540
ttctctntcn ctttcgnncn tnnttnctnc tcttgtccct acctgtcccn ntcatacctt 600
ttcnnaatcg ctnctatcnc cgcctatagt ncaattcnnc tncctnctnn attncctacn 660
nccntcctcn ccatcantnc taacctnctn cntnntctct ntctctgtcc tcancntctc 720
gnccnatttc nttttcccn 739

SEQ ID NO: 143
gatgcatgct cgagcggccg ccagtgtgat ggatatctgc agaattcgcc cttgatagat 60
aattgggttc agcatggggg tcaccacagt gtacatcata gccatgacag tgtcctttag 120
agtagaacta ttagctgatg agcataagta gagaccaata acggttccat agaacagtga 180
caccacagac aggtgggagc cacaagtaga gaaggccttg cagataccct tagaagaagg 240
gaccttgagg atggaggaga caattcttgc ataggaccca aggatgagta ggaatgggat 300
gacaagaatg agccctccca tgataaatat cacccattca ttaactcgag tgtcagagaa 360
ggccagcttc agcagagcag acatatcaca gaaaaagtgg gggatcacat tgtctgcaca 420
aaaacacaac ctggccatga gtaaagtgtg taacatggca tggaaggtgg tcagcaccca 480
ggacagcgcc accaggncga gacagagcat ggggctcatg atggcggngt agtgcngggn 540
gangcagatg nccacantag tgntnatagn ccatggtcac angggaggna gctttcaggn 600
ctttnaataa c 611

SEQ ID NO: 144
gcgtgctcga gcggccgcca gtgtgatgga tatctgcaga attcgccctt gttgcgcaaa 60
gagtacatga aggggttaag tgaaggagtg cccactgcat agaagagacc aaagaacttg 120
cccctccctt gggcatacgg atttttgggc tggaggtaga cagcaatgac tgagctgtag 180
aagagggtga ccacagtgag atgggaggag caggtcccaa aggcctttct ccatgctgtg 240
gnagagttaa tcctcagcac tgnctgggca gtggctccat aagaggcang gatgaggctg 300
agaggcacaa ccacngaaga tgacactgta cacagccaac tgtattttat tgnaggnggn 360
atctccacag gngagnccaa tcagntgatg gntcccnccc atttcanaag tcactntatn 420
tnctnttgnc ngncacgang gtcctnnnng agcngttctt gtccnntctt nactatcgnt 480
taccntccct cntccctcnt nttttctttc cncctncctc ttcntttnnc cntntcccnt 540
gtncncntnt atcttcccta ntncntcttt tnntnctntt tngnnncctt cctctntctt 600
tncntccctc tcnancntat cncttgnncc cncccnntnc c 641

SEQ ID NO: 145
ggttgccnnc gnttaggcat tgggccctct agatgcatgc tcgagcggcc gccagtgtga 60
tggatatctg cagaattcgc ccttccgatg tatttgtttc taagcaacct ctccttcctg 120
gagatttggt ataccacagc agcagtgccc aaagcactgg ccatcctact ggggagaagt 180
cagaccatat catttacaag ctgtcttttg cagatgtact ttgttttctc attaggctgc 240
acagagtact tcctcctggc agccatggct tatgaccgct gtcttgccat cctgctatcc 300
tttacactac ggagccatca tgagtagccn tgctctcagc tgcagctggn cctgggctcc 360
tggntgngct ggtttctcgc cctattnttn ncnnnacnnn ccntantcng ncnctnctct 420

ctttcttnnt tccctttncc tcactcatnc ctcnctctct tttnntgtcc tcttnataac 480
nttgtntntc gnttctcccn ntcntnnctt ctctnttgct tcnctctcct cntttcgnat 540
ccctttgntc tctacnctct tncgnantca ctnnnatntc tnntcacgng cntcctcnnn 600
gatnttcncc tncttactgc tactctctnc tatactnnnc ttntntncat anttcgtctg 660
ctnacnantc tntcactcnt tcccannenn tcnctgtcnt ctgactctcn cctcntntnt 720
nntncctcac cnnntacatg gttccttnnn ntccatctcg tcnntctctc cnntatacgn 780
ttncatactc nctaacttct ctccatcatc ntcacctntc tttctttntc cctngnc 837

SEQ ID NO: 146
gatgatgctc gagcgncgca gtgtgatgga tatctgcaga attcgccctt ccaatgtatt 60
tatttctagg caccactgac ttcttcctct tggccgtcat gtctctggat cgttacctgg 120
caatctgccg accactccgc tatgagaccc tgatgaatgg ccatgtctgt tcccaactag 180
tgctggcctc ctggctagct ggattcctct gggtcctttg ccccactgtc ctcatggcca 240
gcctgccttt ctgtggcccc aatggtattg accacttctt tcgtgacagt tggcccttgc 300
tcaggctttc ttgtggggac acccacctgc tgaaactggn ggctttcatg ctctctacgt 360
tggtggtact gggcccacng gctctgacct cagntttcta ngcccgcatt cttgccactg 420
ttctnagngc ccncnanngc ttgccngagc gaagcanaag atnnttttca cattgcgcac 480
tcggaantta aagggggtgg cgcnncancn nctgggnngc ttcattctnt ctttttactt 540
tnccanngnn tnntngctca ntccctntnc tcntcncaat cntnnnggcn ctcntgntnn 600
gtanactgcc nttaattnga ccnctttccc nacncncac 639

SEQ ID NO: 147
catagatgca tgctcgagcg gccgcagtgt gatggatatc tgcagaattc gcccttccga 60
tgtaagttct ttctaggcac cactgacttc ttcctcttgg ccgtcatgtc tctggatcgt 120
tacctggcaa tctgccgacc actccgctat gagaccctga tgaatggcca tgtctgttcc 180
caactagtgc tggcctcctg gctagctgga ttcctctggg tcctttgccc cactgtcctc 240
atggccagcc tgcctttctg tggccccaat ggtattgacc acttctttcg tgacagttgg 300
cccttgctca ggctttcttg tggggacacc cacctgctga aactggnggc tttcatgctc 360
tctacgttgg tgttactggg ctcactggct ctgacctcag nttcntangc ctgcattctt 420
gtcactgtct caggncccct nnagntgctg ngcgaaggaa agcgcntttc acttgcgcct 480
cnatcttaca ggggtggcat catctnangg ggngnntgca tccttnncta nntncncagg 540
tcccagctat antccaaagt nctnaaaaca nganсctcgg nangannnct nntattctac 600
ccttcttcgt aacctncc 618

SEQ ID NO: 148
cntagatgca ngctcgagcg ggcgccagcg tgnngnanat ctgcagaatt cgcccttcca 60
atgtattttt tctcactaac ttgtctttcc tagatctctg cttcaccacc agttctatcc 120
cccagctgct tttcaatcta ggcagcccag gcaagactat cagccacacg ggctgtgcca 180
tccagctctt catgttcctg ggcctgggtg gcaagagtgt attctcttgg cagccgtggc 240
ctatgaccgc ttcattgcaa tctgcaagcc ccttcactat tctgtcatta tgcaccctca 300
gctgtgctgg aagttggtgt ctgtggcccg gggtgttgg actccncagt tntctaggta 360
tgcctcctgn gactatgaag cttgtcacga tgcggaagat gtaagnttgc ancnttnccn 420
ttnntgngat gccngcntcn tataaaaanc annctgggcg ggtcacagtg cttnngnata 480
gcattnngtc nccttnatnn catcnnattt gcctngnngt ccctcgttcc cantntncan 540
tcnttcntng gcttancntt ctncaccngn ncttncntan ctactccntn ttnnttcntc 600
cttctanctc tncatctttc ttnccntcca tcc 633

SEQ ID NO: 149
gatgcatgct cgagcggccg ccagtgtgat ggatatctgc agaattcgcc cttgttccta 60
agactataca tgaatgggtt tagcatcggg ttgaaagaac tgtaaaatag aaaaaggacc 120
ttctgctgct cctcaggatg gcgggactta ggggccatgt acatgacgat ggcgctgcca 180
aagaagagtc ccactacgca gaggtgggag gagcaggtgg agaaggcctt tctgcggccc 240
tccccagact ggatcctcag gatggccgcc aggatgtgtg agtaggagac cagcaccagg 300

cagagtggtc ccaccaggat gaacatgcag gctgcaaaga tgaccacctg gttgagccag 360
gtatcagcac aggccagcct gaggacagac aggatttcac aagaagaagt ggttgatttc 420
acgaggccca canaaagggc agtcttagga tgaggntcac atggaccata gccaggaggg 480
agccacattg tcccaggaag ngntgnccag agtgatgcag acttttcagg tcntgatgat 540
ngnnttattc ggagagnntg nnagacnggt cancgttccc gntcgtagga caattancac 600
ccancnggng ccttcantna tgtc 624

SEQ ID NO: 150
gatgcatgct cgagcggccg ccagtgtgat ggatatctgc agaattcgcc cttccaatgt 60
atttatttct ctctgacctc tccttcttgg acctctgctt taccacaagt tgtgtccccc 120
agatgctggt caacctctgg ggcccaaaga agaccatcag cttcctggga tgctctgtcc 180
agctcttcat cttcctgtcc ctggggacca ctgagtgcat cctcctgaca gtgatggcct 240
ttgaccgata cgtggctgtc tgccagcccc tccactatgc caccatcatc cacccccgcc 300
tgtgctggca gctggcatct gtggcctggg ttatgagtct ggttcaatcg atagtccaga 360
catcatccac cctccacttg cccttctgtc cccaccagca gatagatgac tttttatgtg 420
aggtcccatc tctgattcga ctctcctgng gagatacctc ctacaatgaa atccagttgn 480
ctgtgtccag tgtcatcttt ggtggntgtg cctctcagcc tcatccttgc ctcttatgga 540
gccactgccc aggcnggggc tgaggattaa ctttgccnna gccatggaag aaaggtcttt 600
nggacctngn n 611

SEQ ID NO: 151
gatgcatgct cgagcggccg ccagtgtgat ggatatctgc agaattcgcc ctttctttat 60
ttcgaagagt atacactagt ggattgaaga gaaacaaata cataggaagg gcgaattcca 120
gcacactggc ggccgttact agtggatccg agctcggtac caagcttgat gcatagcttg 180
agtattctaa cgcgtcacct aaatagcttg gcgtaatcat ggtcatagct gtttcctgtg 240
tgaaattgtt atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa 300
gcctggggtg cctaatgagt gagctaactc acattaattg cgttgcgctc actgtccgct 360
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcgngnaga 420
ggccggnttg cgtattgggc gctcttccgc ttctcgctca ctgactcgct gcgctcggga 480
cgtccggctg cggcgagcgg tatcagctta ctcaanggcc gtantacggt tattcncagg 540
aatnnggggt taacgccngg naaagaacat tgtgngccan angncaagcn taatgcccag 600
gaaccgntan aacgntccc 619

SEQ ID NO: 152
ctcgagcggc gcagtgtgat ggatatctgc agaattcgcc cttcctatgt attatttctc 60
cataatttat ctattgccga tatctgcttc tcttccatca cagcgcccaa ggttctggcg 120
gaccttctgt ctgaaagana gaccatctcc ttcaatcatt gctccactca gatgtttcta 180
ttccacctta ttggaggggc ggntgtatnt nnnccntgtt ncccnatgcg cctnctttc 240
ccntntcntt tcnantcttt ncgcctcctc tcatgcnnnc ccttccntct tattcntgtc 300
gnaatacgct ntctccgnct nctgtctgct catccttgct gttncgtntn canctcatcg 360
ctgtctgtcg tacctnttnc ntnctgtgtc tgcgngntca tncacnntct caancgtctn 420
ccctcactnc tctttncntg ctcttctntn cnccgtgtct tancttcttg ccctgntacg 480
ncncgccgct catatncgng tnctggtatc ccctctnatn ttnttcntcn cctcntntnc 540
cntctcacnn acttcctngt ctctctccan ncttcgacnn ctcnctnatc tccacnacgc 600
acttntctnt ctatatccgc tcttaccgct ctcnncnnan cacncttnnc tctgcatatc 660
agntnntctc ncacncncat nttcttccta cncttctcnc tgtcncacag atctntcnct 720
nctctgctct cgttgntccc cctgncactn cgcaatcnca catatncgtc tctcttcnnt 780
cgccacttat ntngcanctt tctctgcgtt nctctncgat ntccctccnc nntctcncnn 840
ctnatnatcg nttattcnaa tcatactccg tactgtttct gtnctcttnt cntgncnnct 900
agcttctctc tattcantct acnttctntt cgctntctat ccacnctctt cactcncct 959

<220>
<223> Synthetic construct

<221> VARIANT
<222> (1)...(329)
<223> Xaa = Any Amino Acid

<400> 1354
Glu Glu Ile Leu Xaa Ile Ile Ser Gln His Val His Thr Gly Cys Val
1 5 10 15
Gln Asn Xaa Glu Leu Gln Pro Ile Leu Phe Gly Leu Phe Leu Ser Met
20 25 30
Cys Leu Val Met Val Leu Gly Asn Leu Leu Ile Ile Leu Ala Val Ser
35 40 45
Ser Asp Ser His Leu His Thr Pro Thr Tyr Phe Phe Leu Ser Asn Leu
50 55 60
Ser Leu Ala Asp Ile Gly Phe Pro Ser Thr Thr Val Pro Lys Met Ile
65 70 75 80
Val Asp Ile Gln Ser His Ser Arg Val Ile Ser Tyr Ala Gly Cys Leu
85 90 95
Thr Gln Ile Ser Leu Phe Ala Val Phe Gly Cys Met Glu Asp Met Leu
100 105 110
Leu Ser Val Met Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu
115 120 125
Asp Tyr Pro Val Ile Met Asn Pro Cys Phe Cys Gly Phe Leu Val Leu
130 135 140
Leu Ser Phe Phe Leu Ser Leu Leu Asp Phe Gln Leu His Asn Trp Ile
145 150 155 160
Ala Leu Gln Ile Thr Cys Phe Lys Asp Val Glu Ile Pro Ser Phe Phe
165 170 175
Cys Asp Pro Ser Gln Leu Pro His Leu Ala Cys Cys Asp Thr Phe Thr
180 185 190
Asn Asn Ile Val Met Tyr Phe Leu Ala Ala Ile Leu Gly Phe Leu Pro
195 200 205
Ile Ser Gly Ile Phe Tyr Ser Tyr Tyr Lys Ile Val Ser Ser Ile Leu
210 215 220
Lys Val Ser Ser Ser Gly Gly Lys Tyr Lys Ala Phe Ser Thr Cys Gly
225 230 235 240
Ser His Leu Ser Val Val Cys Leu Phe Tyr Gly Thr Ala Leu Gly Gly
245 250 255
Tyr Leu Ser Ser Asp Met Ser Ser Tyr Pro Arg Lys Gly Ala Val Ala
260 265 270
Ser Val Met Tyr Thr Val Val Ala Pro Met Leu Asn Pro Phe Ile Tyr
275 280 285
Ser Leu Arg Lys Arg Asp Ile Lys Ser Ala Leu Gln Gln Leu His Gly
290 295 300
Arg Ile Val Xaa Ser His Asp Leu Ile Ile Gly Ser Ile Leu Xaa Pro
305 310 315 320
Trp Val Gly Lys Gly Ser Lys Val Lys
325

<210> 1355
<211> 321
<212> PRT
<213> Unknown (H38g272 protein)

<220>
<223> Synthetic construct

<400> 1355

Met Glu Ser Pro Asn His Thr Asp Val Asp Pro Ser Val Phe Phe Leu
1 5 10 15
Leu Gly Ile Pro Gly Leu Glu Gln Phe His Leu Trp Leu Ser Leu Pro
20 25 30
Val Cys Gly Leu Gly Thr Ala Thr Ile Val Gly Asn Ile Thr Ile Leu
35 40 45
Val Val Val Ala Thr Glu Pro Val Leu His Lys Pro Val Tyr Leu Phe
50 55 60
Leu Cys Met Leu Ser Thr Ile Asp Leu Ala Ala Ser Val Ser Thr Val
65 70 75 80
Pro Lys Leu Leu Ala Ile Phe Trp Cys Gly Ala Gly His Ile Ser Ala
85 90 95
Ser Ala Cys Leu Ala Gln Met Phe Phe Ile His Ala Phe Cys Met Met
100 105 110
Glu Ser Thr Val Leu Leu Ala Met Ala Phe Asp Arg Tyr Val Ala Ile
115 120 125
Cys His Pro Leu Arg Tyr Ala Thr Ile Leu Thr Asp Thr Ile Ile Ala
130 135 140
His Ile Gly Val Ala Ala Val Val Arg Gly Ser Leu Leu Met Leu Pro
145 150 155 160
Cys Pro Phe Leu Ile Gly Arg Leu Asn Phe Cys Gln Ser His Val Ile
165 170 175
Leu His Thr Tyr Cys Glu His Met Ala Val Val Lys Leu Ala Cys Gly
180 185 190
Asp Thr Arg Pro Asn Arg Val Tyr Gly Leu Thr Ala Ala Leu Leu Val
195 200 205
Ile Gly Val Asp Leu Phe Cys Ile Gly Leu Ser Tyr Ala Leu Ser Ala
210 215 220
Gln Ala Val Leu Arg Leu Ser Ser His Glu Ala Arg Ser Lys Ala Leu
225 230 235 240
Gly Thr Cys Gly Ser His Val Cys Val Ile Leu Ile Ser Tyr Thr Pro
245 250 255
Ala Leu Phe Ser Phe Phe Thr His Arg Phe Gly His His Val Pro Val
260 265 270
His Ile His Ile Leu Leu Ala Asn Val Tyr Leu Leu Leu Pro Pro Ala
275 280 285
Leu Asn Pro Val Val Tyr Gly Val Lys Thr Lys Gln Ile Arg Lys Arg
290 295 300
Val Val Arg Val Phe Gln Ser Gly Gln Gly Met Gly Ile Lys Ala Ser
305 310 315 320
Glu

<210> 1356
<211> 327
<212> PRT
<213> Unknown (H38g273 protein)

<220>
<223> Synthetic construct

<221> VARIANT
<222> (1)...(327)
<223> Xaa = Any Amino Acid

<400> 1356
Met Thr Trp Ser Gly Gly Thr Ile Val Gly Glu Xaa Gly Glu Phe Val
1 5 10 15
Leu Leu Gly Phe Pro Ala Pro Ala Pro Leu Gln Val Leu Leu Phe Ala
20 25 30
Leu Leu Leu Leu Ala Tyr Val Leu Val Leu Thr Glu Asn Thr Leu Ile

<400> 423
atggaatctc ctaatcacac tgatgttgac ccttctgtct tcttcctcct gggcatccca 60
ggtctggaac aatttcattt gtggctctca ctccctgtgt gtggcttagg cacagccaca 120
attgtgggca atataactat tctggttgtt gttgccactg aaccagtctt gcacaagcct 180
gtgtaccttt ttctgtgcat gctctcaacc atcgacttgg ctgcctctgt ctccacagtt 240
cccaagctac tggctatctt ctggtgtgga gccggacata tatctgcctc tgcctgcctg 300
gcacagatgt tcttcattca tgccttctgc atgatggagt ccactgtgct actggccatg 360
gcctttgatc gctacgtggc catctgccac ccactccgct atgccacaat cctcactgac 420
accatcattg cccacatagg ggtggcagct gtagtgcgag gctccctgct catgctccca 480
tgtcccttcc ttattgggcg tttgaacttc tgccaaagcc atgtgatcct acacacgtac 540
tgtgagcaca tggctgtggt gaagctggcc tgtggagaca ccaggcctaa ccgtgtgtat 600
gggctgacag ctgcactgtt ggtcattggg gttgacttgt tttgcattgg tctctcctat 660
gccctaagtg cacaagctgt ccttcgcctc tcatcccatg aagctcggtc caaggcccta 720
gggacctgtg gttcccatgt ctgtgtcatc ctcatctctt atacaccagc cctcttctcc 780
tttttacac accgctttgg ccatcacgtt ccagtccata ttcacattct tttggccaat 840
gtttatctgc ttttgccacc tgctcttaat cctgtggtat atggagttaa gaccaaacag 900
atccgtaaaa gagttgtcag ggtgtttcaa agtgggcagg gaatgggcat caaggcatct 960
gag 963

<210> 424
<211> 982
<212> DNA
<213> Unknown (H38g273 nucleotide)

<220>
<223> Synthetic construct

<400> 424
atgacatgga gtggcggaac catagtggga gagtgagtga gtttgtgttg ctggggcttc 60
cctgctcctg cgccactaca ggtactattg tttgcccttt tgctgctggc ctatgtgttg 120
gtgctgactg agaacacact catcattatg gcaattagga accattccac cctccacaaa 180
cccatgtact tttttctagc taatatgtcc tttctggaga tctggtatgt cactgtcact 240
attcccaaga tgcttgctgg ctttgttgga tccaaacagg atcatggaca gctaatctcc 300
tttgagggat gcatgacaca gctctacttt ttccttggct tgggctgcac tgagtgtgtc 360
cttctcgctg ttatggccta tgatcgctat atggccatct gctatcctct ccactaccca 420
gtcattgtca gtggccggct gtgtgtgcag atggctgctg gctcttgggc tggaggtttt 480
ggcatctcca tggtcaaagt tttttcttatt tctggcctct cttactgtgg ccccaacatc 540
atcaaccact ttttctgtga tgtctctcca ttgctcaacc tctcatgcac tgatatgtcc 600
acagcagagc ttacagattt catcctggcc atttttattc ttctagggcc actctctgtc 660
actggggcct cctatgtggc cattactggt gctgtgatgc acatacctte ggctgctgga 720
cgctataagg ccttttccac ctgtgcctct catctcactg ttgtgataat cttctatgca 780
gccagtatct tcatctatgc tcggccaaag gcactctcag cttttgacac caacaagttg 840
gtctctgtac tgtatgctgt cattgtacca ttgctcaatc ccatcattta ctgcctgcgc 900
aatcaagagg tcaagagagc cctatgctgt actctgcacc ctgtaccagc accaggatcc 960
tgaccccaag aaagctagca ga 982

<210> 425
<211> 936
<212> DNA
<213> Unknown (H38g274 nucleotide)

<220>
<223> Synthetic construct

<400> 425
atggaagcag gaaaccaaac aggattttta gagtttatcc ttctcggact ctctgaggat 60
ccagaactac agccgttcat atttgggctg ttcctgtcca tgtacctggt gacggtgctg 120
ggaaacctgc tcatcatcct ggccatcagc tctgactccc acctccacac ccccatgtac 180
ttcttcctct ccaacctgtc ctgggttgac atctgtttca gcacttgcat cgtccccaag 240
atgctggtga acatccagac cgagaacaaa gccatctcct acatggactg cctcacacag 300
gtctatttct ccatgttttt tcctattctg gacacgctac tcctgaccgt gatggcctat 360